University of Groningen

The effects of sampling frame designs on nonresponse and coverage error

Kölln, Ann-Kristin; Ongena, Yfke P.; Aarts, Kees

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

THE EFFECTS OF SAMPLING FRAME DESIGNS ON NONRESPONSE AND COVERAGE ERROR: EVIDENCE FROM THE NETHERLANDS

ANN-KRISTIN KÖLLN*
YFKE P. ONGENA
KEES AARTS

Survey researchers sometimes face several options to formally define and draw random probability samples from their target population of individuals. Relatively little is known about the consequences for the quality of survey estimates when different sampling frames are used, for example frames that list addresses versus individuals While the initial choice for a sampling frame design may be based on diverse criteria, we argue that any decision has consequences for the quality of survey estimates. We hypothesize that knowing respondents' names upfront decreases nonresponse and coverage error. This can be accomplished by either using person-based sampling frames or augmenting nonperson-based sampling frames with names. We systematically compare sampling frame designs in the context of face-to-face surveys in connection with the European Social Survey (ESS) with the help of three quasi-experimental datasets from a single country. Even the most conservative measures support our hypothesis that the presence of names in the sampling frames could improve response rates, noncontact rates, cooperation rates, and ineligibility rates by between 2.5 and 6 percentage points. Additionally, the accuracy of population estimates could increase. These results suggest that survey researchers collecting individual-level data would be best advised to use well-maintained sampling frame designs (augmented) with person-specific information.

ANN-KRISTIN KÖLLN is with the Department of Political Science, Aarhus University. YFKE ONGENA is with Department of Communication Sciences, University of Groningen. KEES AARTS is a member of faculty of Behavioural and Social Sciences, University of Groningen.

*Address correspondence to Ann-Kristin Kölln; Department of Political Science, Aarhus University, Bartholins Allé 7, 8000 Aarhus, Denmark E-mail: koelln@ps.au.dk.

KEYWORDS: European Social Survey; Sampling; Sampling frame designs; Survey estimate quality.

1. INTRODUCTION

Studies in survey methodology acknowledge that the quality of survey estimates is a consequence of choices made during the entire sampling process. The sampling frame design is one important part of this process. The sampling frame design is the ordering characteristic (e.g., names, addresses, or phone numbers) of elements of the target population on a list from which the sample is drawn. The literature distinguishes between four types of sampling frame designs. Lists can contain telephone numbers (telephone-based sampling), names of persons (person-based sampling), addresses (address-based or household-based sampling), or geographical units (area-based sampling). An emerging literature stresses the importance of systematic comparisons between sampling frame designs (Link, Battaglia, Frankel, Osborn, and Mokdad 2006, 2008; Tsuchiya and Synodinos 2015). This is motivated by the following points: (1) The recent rise in individual-level data collection by governments and businesses has increased the number of options to identify target populations, such as through e-government initiatives[1] or personal information given to online businesses; (2) legal regulations on data protection have increasingly limited the scope of available sampling frame designs in some countries; (3) Surveys of probability samples are experiencing decreasing response rates. All three developments call for more systematic comparisons between sampling frame designs and the resulting quality of survey estimates.

The goal of any realized survey sample is to generalize to the target population. This requires high-quality survey estimates and minimizing the total survey error. The quality of survey estimates are affected by measurement error and representation error. There is a widely held belief that up-to-date, person-based registries are best at minimizing these errors (see Harter, Battaglia, Buskirk, Dillman, English, et al. 2016). However, since such sampling frames are not always available for the general population, literature assessing this claim is scarce.

This paper takes advantage of a rare opportunity that several sampling frame designs are available for a target population of individuals aged fifteen and over residing in private households. It studies to what extent commonly used sampling frame designs for face-to-face interviews in the social sciences affect the quality of survey estimates. In particular, we focus on the differences between person-based and address-based sampling frame designs and their effects

---

1. For example, as of 2017, residents and businesses in the Netherlands must be able to conduct all of their business, such as applying for permits or objecting to decisions, with the government online. For more information, please visit: https://www.houseofrepresentatives.nl/dossiers/digital-government-2017-project [last accessed 08/22/2017].

on (unit) nonresponse error and coverage error. We define nonresponse as sampled individuals who cannot be reached or who do not want to or cannot participate. Coverage error occurs if some individuals in the population are not included in the sampling frame (undercoverage) or if the sampling frame covers too many (overcoverage) (Groves, Fowler, Couper, Lepkowski, Singer, et al. 2009).

The initial choice for a sampling frame design may be based on diverse criteria (see also Särndal, Swensson, and Wretman 2003; DiGaetano 2013), such as the following:

– *Availability*: Researchers may not always have access to all possible sampling frame designs for the target populations.
– *Ability to find and contact all units*: Sampling frame designs vary in their ability to allow researchers to locate and to contact units. Some only offer a location on the map, while others offer names, addresses, or phone numbers. It means that the ability to contact the unit is distinct from the ability to locate it.
– *Organization of units:* If units on a frame are organized by size or geography, this will simplify selection procedures.
– *Interviewing methodology*: Researchers' choice of interviewing methodology may make some sampling frame designs more attractive than others. For example, in CATI surveys, telephone-based sampling is a more obvious choice than other sampling frames because it provides a higher rate of (accurate) phone numbers.
– *Coverage*: Sampling frame designs vary as to their coverage of the target population, and thus, in the share of eligible units, the ability to identify ineligible units, and the duplication or clustering of units.
– *Accuracy of (contact) information*: Sampling frames may be out of date, resulting in incorrect information that increases nonresponse and the fieldwork duration.
– *Availability of auxiliary information and ability to match data frame with them*: Some frames include not only contact information but also data such as gender, age, and ethnicity. When absent, some frames may be matched with other frames to include auxiliary information. Such information may be used for stratification and weighting or for determining ineligibility.
– *Costs*: Some frames are costly to make available, and others add to the screening costs because of duplication and ineligibility rates.

The choice of a sampling frame design has consequences for nonresponse and coverage error. Specifically, we show that a person-level sampling frame design (or one that is augmented with individual-level information) gives access to respondents' names, with two positive consequences. First, our results suggest that a frame using names increases response and cooperation rates and decreases noncontact rates because households for which names are documented in address databases are less concerned about providing information and can be addressed directly by name. Second, our results suggest that a sampling frame design with up-to-date information in which frame elements match

target elements minimizes threats to coverage error because they encompass more detailed information on the target elements. This means that the quality of survey estimates from a target population of individuals should be highest if a well-maintained sampling frame design (augmented) with person-specific information is used. Alternative sampling frame designs using characteristics such as addresses, households, or areas increase threats to full coverage and should have negative implications for the quality of survey estimates, provided that individuals are the target population.

Using the rare instance of several available sampling frame designs, this paper examines the extent to which different sampling frame designs affect the resultant survey estimates. Specifically, we compared two versions of an address-based sampling frame with a person-based sampling frame to assess the impact of different sampling frame designs on response error and coverage error. We motivate this unique quasi-experimental design by showing that cross-country comparisons of paradata from the European Social Survey (ESS), such as response or noncontact rates, are heavily constrained in their conclusions. Our design overcomes some of the drawbacks inherent to analyzing paradata with country-specific effects. An in-depth analysis of differences in sampling frame designs applied to one specific case, the Netherlands, provides support for our argument that differences between sampling frame designs have important implications for the degree of nonresponse and coverage error. We use the ESS throughout this study as a running example representative of large, cross-country comparative surveys in the social sciences.

2. SAMPLING FRAME DESIGNS AND THE QUALITY OF SURVEY ESTIMATES

2.1 Types of Sampling Frame Designs

The ESS gives national coordinators freedom to decide on the design that fits the country best. It assumes that several sampling frame designs are equivalent in the quality of data they produce (see European Social Survey 2012).

Our analysis of sampling frame designs employed during most recent rounds of ESS data collection between 2002 and 2012 shows strong patterns of over-time continuity. Of the thirty-four countries that participated more than once, 61.8 percent used the same sampling frame design throughout. An additional 26.5 percent switched the sampling frame design only once and a further 11.8 percent twice. This may indicate either that most national coordinators only had one option for a sampling frame design or, alternatively, that they chose the same sampling frame design based on the previously mentioned selection criteria. According to information from personal contact with some national coordinators (Belgium, Hungary, Spain), the decision to switch was based on accessibility and coverage. However, we argue that irrespective of

the motivating reasons, any decision (even those with limited options) has consequences for survey sample quality because of the intrinsic differences between sampling frame designs.

In general, the literature distinguishes between four types of sampling frame designs. Lists can contain telephone numbers, names of persons, addresses, or geographical units. Table 1 summarizes the different, theoretically possible methods. We recognize that not all methods are always accessible to researchers, and we also acknowledge the variability in quality between different methods. We assume for our discussion that all available sampling frames are of the same quality, but we later relax this assumption when making more specific recommendations.

One of the major differences between the designs lies in their mode of first contact. For telephone-based sampling (TBS), the phone is necessarily the first contact, while the other designs are more flexible. However, based on the typical availability of auxiliary information, administration by telephone is only less likely in address-based sampling (ABS) and in area-based sampling (ARBS). Farrell and Petersen (2010, p. 114) argue that in recent years, the simultaneous decrease in landlines and respondents' increased skepticism toward unsolicited calls have limited the efficacy of TBS while increasing its costs. It makes the search for alternative sampling frames and their comparison all the more important (see also Harter et al. 2016, p. 1–1). The following

Table 1. Overview of Different Sampling Frame Designs

Design Of sampling	Description	Mode of contact	Element on list
Telephone-based sampling (TBS)	Usually random selection of numbers (RDD) from a set of all possible four-digit numbers within existing telephone exchanges	phone	Household (possibly individual if mobile phone)
Person-based sampling (PBS)	Individuals are selected from the registry	phone/mail/ personal visit	Individual
Address-based sampling (ABS or household-based sampling (HHBS)	Addresses are selected from the registry or household list	phone/mail/ personal visit	Address/ Household
Area-based sampling (ARBS)	For each area, an inventory of households is made from which households are sampled	phone/mail/ personal visit	Geographical units

NOTE.—If no sampling frame is available, random route sampling is used.

discussion will focus mostly on the other designs and their differential impact on survey quality.

Importantly for our argument, a second major difference between the designs pertains to the level of person-specific information available to researchers. Person-based sampling (PBS) designs include person-specific information such as name, age, and gender next to the address. They are often based on government registries, but this is not a necessary condition. Household-based sampling (HHBS) and ABS designs only include addresses of residential homes but no information on the individual(s) living under the address (although in the case of HHBS, additional information is sometimes retrievable). While HHBS is based on consumer registers such as electricity bills, ABS tends to encompass a wider range of addresses. These addresses can be augmented with names based on databases that are publicly or commercially available. Area-based sampling designs are a hybrid since they cover geographical units, but during the multi-stage sampling process, lists of addresses and sometimes names of residents are added (see Iannacchione 2011). This brief overview suggests that sampling frame designs differ in their level of abstraction and thus encompass more or less detailed information on individuals.

2.2 Consequences of Information Differences in Sampling Frame Designs

Information differences between sampling frame designs may yield several consequences for the quality of survey estimates, particularly for the degree of nonresponse and coverage error. In this section, we discuss them and derive our specific hypotheses related to nonresponse and coverage error from existing literature. Our overall hypothesis is that sampling frame designs containing personal information on individuals, such as names, decrease the degree of nonresponse and coverage error. This should be particularly true when information on sampling frames is current.

Link et al. (2006, 2008) and Link and Lai (2011) were the first to show in several studies that an ABS design performed better than a TBS design on important measures such as response rates, coverage, and costs. Similarly, Tsuchiya and Synodinos (2015) show with experimental data from Japan that while an ARBS design achieved a response rate of 18 percent, the PBS design yielded a response rate of 28.7 percent. They do not find any differences in participating respondents' demographics or expressed opinions. The authors speculate that the difference in response rates may be the result of using respondents' names in the PBS design. This suggests that respondents feel more personally involved when addressed personally and are thus more inclined to participate in the survey. However, differences in sampling frame designs cannot always be easily distinguished from differences occurring from varying first contact and interviewing methods. ABS designs are not

necessarily less personal than PBS designs, for example, provided that the respondent's name is known. Still, it may be difficult to obtain accurate information about respondents' names for ABS designs in different countries and target populations. Therefore, sampling frame designs that allow for addressing the individual personally should increase response rates and cooperation while lowering noncontact rates.[2] This leads to the first hypothesis:

> H1: Samples drawn from PBS frames show higher response rates compared with samples drawn from ABS.

This hypothesis can be explained by two subordinate hypotheses:

> H1a: Samples drawn from PBS frames show higher cooperation rates compared with samples drawn from ABS.

> H1b: Samples drawn from PBS frames show lower noncontact rates compared with samples drawn from ABS.

Different sampling frame designs can also affect the quality of survey estimates because of their varying threats to full coverage. A close match between frame elements and target elements increases the probability that the sample can meet the basic requirement of a random probability sample in which each element has a measurable (and preferably equal) chance of selection (AAPOR 2014). Coverage error can impede this basic requirement if the frame lists too many elements (overcoverage), does not list all elements (undercoverage), lists some elements more than once (duplication), or if some elements are nested in higher units (clustering) (Groves et al. 2009). Each of these four types can lead to coverage error, but here, we focus on overcoverage and undercoverage because duplication and clustering are special cases of overcoverage and undercoverage, respectively. We argue that to the extent that sampling frame designs differ in their levels of abstraction (e.g., listing households versus listing individuals), some should be better at minimizing coverage errors than others.

Overcoverage occurs when a sampling frame covers elements that the target population does not include. This includes ineligible units because they are not part of the target population. For instance, in the case of the ESS, a frame could also list children below the age of fifteen. Ineligibility can often only be identified during the data collection process, but it is also possible (especially in self-administered surveys) that ineligibility of a unit is never detected. Availability of individual-level information can reduce this problem because it allows researchers to draw a sample from the precise target population after ineligibles are eliminated (see Groves et al. 2009). If only the share of ineligibles on a frame were known instead of the precise elements, more elements could be sampled to adjust for the dropout of units (Groves et al. 2009; European Social

---

2. Definitions of noncontacts vary by sampling frame design. Yet, even if it is defined at the household level, addressing individuals within that household personally should improve noncontact rates.

Survey 2012). But knowledge about the share of ineligibles also presupposes individual-level information. However, if not even the share of ineligibles is known, gross sample sizes cannot be adjusted, and the net sample size will be lower, which reduces statistical power. Therefore, a small share of ineligibles is not directly related to survey quality, but it can be an indicator for the coverage error, as Groves et al. (2009, p. 93) indicate. While the availability of names does not necessarily mean that other background information of individuals is known, it nonetheless increases the likelihood. Researchers who have access to individuals' names might also have access to other background information and should therefore be able to compose a sampling frame that reduces the number of ineligibles. This suggests that for a target population of individuals, a PBS design in which names and other individual-level information are known should reduce the share of ineligibles. This leads to second hypothesis, as stated below.

H2: In a context of a target population of individuals, samples drawn from PBS frames have fewer ineligibles and thus lower overcoverage compared with samples drawn from ABS.

Several studies also show how problems of undercoverage can arise with ABS/HHBS, but in particular with ARBS (Iannacchione 2011; Kalton, Kali, and Sigman 2014; Koch, Halbherr, Stoop, and Kappelhof 2014). This mostly occurs when the target population consists of individuals, but the sampling units are higher-level units, such as addresses or households. In such instances, a second smaller frame per address is required that lists all individuals living under that address (Groves et al. 2009). This can be a source of undercoverage because the second, smaller frame may not list all individuals officially residing under that address (see Tourangeau, Shapiro, Kearney, and Ernst 1997; Kalton et al. 2014). Another threat to ARBS and possibly ABS is missing sample units from first enumeration (Eckman and Kreuter 2011). To the extent that residents of these missed units do not represent a random subsample of the target population, this can be another important source of undercoverage.

Koch et al. (2014), for example, compared estimates of variables (e.g., gender, age, marital status, work status, nationality, household size) from the ESS 2010 with estimates from another high-quality European survey that employs different sampling frame designs—the 2010 European Union Labour Force Survey (LFS). They found that the average difference between ESS sample estimates and LFS estimates was twice the size when the frames of the samples were based on addresses or households, compared with individuals (Koch et al. 2014, p. 19). These initial results suggest that a PBS frame could improve the estimates of subpopulation statistics because of a lower risk of coverage error. The importance of lowering the risk also depends on the substantive interest of the survey (see Groves et al. 2009, p. 54). For a general survey, subpopulation statistics are also important. Based on these initial findings, we propose the exploratory hypothesis that the size of subgroups of the population

could be systematically misrepresented in non-PBS frames. This leads to the following exploratory third hypothesis.

> H3: In a comparison with population statistics, samples drawn from PBS frames will show group estimates that are closer to the true population values compared with samples drawn from ABS.

This discussion shows that some sampling frame designs are better equipped to minimize biases compared with others when individuals are the target population. These sampling frame designs should reduce nonresponse and coverage error because of their lower level of abstraction and the availability of person-level information. We note that ABS designs augmented with names contain more person-specific information than a normal ABS design but less than a PBS design. This augmented ABS design should therefore range between the latter two in terms of the degree of nonresponse and coverage error.

3. DATA AND METHODS

Paradata offer a good source of data to compare properties of surveys. The ESS employs a "standardized paradata collection" in order "to control and compare the data collection process" in participating countries (Stoop, Matsuo, Koch, and Billiet 2010, p. 407/420). The ESS has taken all possible precautions to facilitate the comparison of data collection processes and their resulting samples. For rounds one through six, we collected information on sampling frame designs for a total of 160 country-year observations from the descriptions included in the Data Documentation Reports of the ESS. We found that thirty-six surveys were based on addresses (ABS), twelve on areas (ARBS), thirty-six on households (HHBS), and seventy-two on persons (PBS). The categorization was based on the first set of units from which random probability samples were obtained (see table A1 in the supplementary data online for information on the countries' classification).

Despite this diversity in sampling frame designs across a single European survey project, these observations are far from being independent, as shown above. It means that analyses based on these paradata will be limited because the selection of sampling frame designs often remained the same within countries and across rounds. Therefore, we test our hypotheses with a quasi-experiment from a single country that promises high internal validity.

We assess the impact of switching sampling frame designs on nonresponse error and coverage error when several well-maintained sampling frame designs are available for a target population of individuals aged fifteen and over. It is important to evaluate what difference switching sampling frame designs make because, theoretically, PBS should be considered the gold standard, but it is unknown how well ABS performs in comparison? After all, in practice,

AAPOR considers ABS frames "the best possible frames for today's household surveys in the United States" (Harter et al. 2016, p. 1–1).

For this analysis, we use the ESS data from the Netherlands. Using only data from the Netherlands holds constant many potential confounders. We first drew a sample from a list ordered by addresses (gross sample = 1,546, net sample = 1,184). Then we attempted to match the 1,564 addresses with a name using the "Nationaal Consumenten Bestand" (NCB). Among the 1,546 addresses, names were matched for 998 addresses; this constitutes the "ABS augmented" sampling design. The rest of the 548 addresses were not able to be matched with a name; they are the "ABS address only" sampling design. The success rate of finding names to corresponding addresses, which refers to any name being matched to the address with no possibility of determining whether the name was correct, was 64 percent (998 out of 1,546). This is comparable to success rates found in earlier studies (i.e., between 66 percent and 78 percent, Link et al. 2008). The technique is common practice, and it was used in previous rounds of the ESS Netherlands. Although this procedure confounds the availability of names and the actual use of names in addressing respondents, effects are most likely due to the mere availability of names irrespective of using the name on the mailing envelope. Therefore, no further segmentation in use of names was made in the quasi-experimental design, and names were used for all cases with matching names in the augmented ABS sample.

The third dataset is a PBS sample (gross sample = 2,710; net sample = 1,677), is drawn from the Dutch population register. Data for the registry are collected by the municipalities. The central registry updates these annually and immediately via an electronic system whenever a demographic event occurs. The registry covers all legal residents in the Netherlands, including (EU-)foreigners; nonpermanent residents are registered in the municipality of The Hague. However, illegal immigrants are not part of the PBS sample. Legal residents have very few incentives to not update their information after they have moved houses. Usually, landlords, electricity, and water supply companies also require a proof of residency in that municipality. However, individuals receiving social benefits could have an incentive not to update their information, but incorrect registration is also subject to an administrative fine of up to €325. When it comes to the level of accuracy of the recorded legal population, a governmental investigation in summer 2009 found that only about 5 percent of cases contained errors (Tweede Kamer der Staten-Generaal 2009). It suggests only a small mismatch between the target population and the frame.

The samples are comparable in a number of important aspects and are summarized in table 2. Fieldwork for all samples took place at the same time (shortly after the Dutch parliamentary elections in autumn 2012). Respondents of all samples were approached with questionnaires on the national election and sociopolitical attitudes in general. Questionnaires and interview lengths

Table 2. Summary of Samples

Sample	Description	Source of data
ABS sample address only ($n = 548$)	Survey data of Dutch individuals, drawn from a list of addresses	Cendris
ABS sample augmented ($n = 998$)	Survey data of Dutch individuals, drawn from a list of addresses and augmented with names	Cendris/Nationaal Consumenten Bestand (NCB)
PBS sample ($n = 2,710$)	Survey data of Dutch individuals drawn from a list of individuals	Statistics Netherlands

were comparable. Interviews with respondents in the ABS samples took on average 44.98 minutes (standard deviation = 17.77 minutes) and, in the PBS sample, 46.81 minutes (standard deviation = 26.55 minutes). The differences are not statistically significant (t [2.273] = 1.565, $p = 0.118$).

All respondents were first mailed an advance notice letter, including a €5 coupon as an incentive and an announcement of the personal visit of an interviewer. Respondents in the ABS sample for which no names were found received an advance notice letter addressed "to the residents of" followed by the street name and house number (a very common way of addressing direct mail in the Netherlands). The advance notice letter sent to the augmented part of the ABS sample and the PBS sample included a specific name. Questions asked in both parts of the ABS sample are identical; the PBS questionnaire differs in details, such as precise question wording or response scales. Interviewers for all samples were instructed to visit the respondent for the first time a few days after the arrival of the introductory letter and to only list a respondent as a "noncontact" after six idle visits.

Both parts of the ABS sample are identical in their fieldwork organization (GfK), the fieldwork period, and interviewers used (seventy-three different interviewers). Assigned interviewers did not differ in their year of birth, gender, or years of experience at GfK. A one-way ANOVA for year of birth (t [1.544] = 1.444, $p = 0.230$) and χ^2 tests for gender ($p = 0.368$) and years of experience in categories ($p = 0.838$) returned no statistically significant results. Interviewers were not aware of the research. This further suggests the comparability of the samples and strengthens our argument for a quasi-experimental design. The PBS sample has been obtained by a different fieldwork organization (Statistics Netherlands), but close attention has been paid to mirror the work for the ABS sample. It could be that differences in fieldwork organizations (commercial versus state-run) confound observed differences, for instance, in cooperation rates or response rates. But no information is available on the extent to which this has happened here or in the past. This project is the first to conduct almost identical surveys by two

fieldwork companies under highly similar conditions. To the extent that confounding occurred, differences between the augmented and nonaugmented ABS samples should be taken more seriously because the fieldwork organization was held constant. Although fieldwork organizations differed, all interviews were conducted face-to-face, and interviewers in both the PBS sample and the augmented ABS sample knew respondents' names before the first visit. Social desirability could show in the reported vote choice for one of the more extreme parties. But we only find small differences in reported vote choice across the samples (PBS = 7.0 percent; augmented ABS = 6.2 percent; nonaugmented ABS = 8.8 percent). These differences and—above all—the large similarities lead to a quasi-experimental design.

To measure response rates, we use AAPOR's RR1 definition and code successfully conducted interviews using the Contact Files. The variable "cooperation" differentiates between a successful interview and no interview (excluding noncontacts). Noncontact and ineligibility are defined as respondents who did not respond to any contacting attempt and those who were not eligible, respectively. For the analyses of response rates, cooperation rates, noncontact, and ineligibility rates, no survey weights are applied. Consequently, the results can not generalize to the larger population, as is generally the case when no additional information on nonrespondents is available.

We measure the degree of undercoverage with the share of specific subgroups of the population as provided in the individual-level data files: women, the low-educated aged fifteen to sixty-four, the unemployed, seniors, and married respondents. We select these groups to include both smaller (the low-educated and unemployed) and larger groups (women, senior citizens, and married individuals) and groups that are more likely (women, the unemployed, and senior citizens) and less likely to answer surveys (the low-educated and married individuals), respectively. Selecting these groups with their different characteristics safeguards us from the risk of a selection bias. Other background characteristics were not available for all survey samples. We estimate these groups' sizes with the population statistics taken from Statistics Netherlands' population register either as an average for the fieldwork period (September to November 2012) or for the entire year 2012 (Statistics Netherlands 2012). The unemployed were defined as those who were part of the working force population in the Netherlands but were without a job. For the survey samples, they were defined as those who responded "no" to the question of being currently employed. This also includes students and pensioners. However, when checked against responses to the survey question, "What was your main activity last week?" only around 10 percent of respondents who initially claimed to be employed indicated having been a student, pensioner, unable to work, or unemployed during the last week. For this part of the analysis, we used weights based on population values from Statistics Netherlands for gender, age, and marital status measured at the time the survey was conducted. They are post-stratification weights using simple cell weighting. Appendix 2 of the

supplementary data online lists the specific weights, the code for the analyses, and the tests of homogeneous proportions.[3]

4. RESULTS

We begin by testing Hypotheses 1 and 2 on the differential impact of sampling frame designs on response rates, noncontact rates, cooperation rates, and ineligibility rates. Table 3 summarizes the achieved values and the associated χ^2 values. Results are in the expected direction: compared with the nonaugmented ABS sample, values from the PBS design show higher response rates (61.9 percent versus 37.2 percent) and cooperation rates (66.9 percent versus 41.9 percent), and lower noncontact rates (4.3 percent versus 7.8 percent) and ineligibility rates (1.45 percent versus 8.7 percent). In other words, the PBS design improved response rates by 24.7 percentage points, cooperation rates by 25 percentage points, noncontact rates by 3.5 percentage points, and ineligibility rates by 7.25 percentage points. All differences are statistically significant.

The last column in table 3 also shows that almost all differences between the augmented version of the ABS sample and the PBS sample are statistically significant. The PBS sample seems to perform better than the augmented ABS sample. Only the difference in noncontact rates is not statistically significant. We can only speculate that the high standard for concluding noncontact (i.e., six idle visits) makes the differences between the augmented ABS and the PBS not statistically significant. While the same standard was applied to the non-augmented sample, differences were statistically significant. Our theory suggests that the absence of names might be responsible for that.

The comparison between the pure ABS data and the augmented version of the ABS provides the most conservative test of our hypotheses, albeit not experimental (column four). These two samples were almost identical, except one was augmented with names. This is only a small difference, and although there are two confounded explaining factors (households less concerned about providing names and the possibility of more personally addressing respondents through the usage of names), both factors are hypothesized to independently favor the augmented sample. The results show that the augmented version of the ABS design performs better than the ordinary ABS design even though it is worse than the PBS design. Here, indicators for nonresponse and coverage error improve by between 2.5 to 6 percentage points for the augmented sampling design (over the "ABS address only" sampling design). Most differences are statistically significant, except for the difference in cooperation rates (41.9 percent versus 46.4 percent, respectively, in the address only and augmented samples).

To substantiate these findings on the differences between ABS address only and ABS augmented, logistic regressions were conducted on each of the

---

3. The results remain mostly the same when no weights are applied.

Table 3. Survey Quality of Augmented and Non-Augmented Parts of ABS Sample and PBS Sample

	ABS sample (address only)	ABS sample (augmented)	PBS sample	χ^2 ABS address only vs augmented (df = 1)	χ^2 ABS address only vs PBS (df = 1)	χ^2 ABS augmented vs PBS (df = 1)
Noncontact rate	7.8 %	5.3 %	4.3 %	3.907*	11.84**	1.52
Response rate	37.2 %	43.2 %	61.9 %	4.890*	105.59**	101.18**
Cooperation rate	41.9 %	46.4 %	66.9 %	2.468	101.76**	117.38**
Ineligibility rate	8.7 %	3.5 %	1.45%	19.207**	93.88**	15.73**

Notes: *$p < 0.05$, ** $p < 0.01$; no survey weights applied.

parameters using the contact files from the address samples. They were fitted to individual case-level data without any weights. The main independent variable distinguishes between instances in which the name was known ($= 1$). We also include all possible confounders available to us: a dummy for heavily urbanized areas, dummy variables for five districts in the Netherlands with the three largest cities (Amsterdam, Rotterdam, The Hague) as the reference category, the number of contact attempts and the interviewer's gender, age, and experience. Detailed tables can be found in the supplementary data online (table A.2). The results suggest that the availability of respondents' names reduces the nonresponse and coverage error. Specifically, it improves the odds of having an interview or achieving cooperation by 27 percent ($p < 0.05$) and 22 percent ($p = 0.100$), respectively, while lowering the odds of no contact at all by 41 percent ($p < 0.05$) and reducing the ineligibility rate even by 63 percent ($p < 0.00$). Even though the effect on cooperation rate just fails to reach statistical significant at the 90-percent level, the combined results still indicate support for hypothesis one and two.

Next, we test hypothesis three pertaining to the impact of sampling frame designs on undercoverage. The comparison of estimated group sizes with true population values (table 4) shows that the estimates regularly match the true population value with 95 percent certainty. In all three samples, four out of five estimates and their confidence intervals cover the true population value. The estimates and confidence intervals *not* covering the true population values are those for 'older than 65' for the ABS address only sample, 'primary school education' for the ABS augmented sample, and 'without work' for the PBS sample.

Table 4. Survey Estimates Across Samples and True Population Values

	ABS address only	ABS augmented	PBS
Female (true value = 50.5 %)			
Estimate	53.2 %	49.2 %	50.4 %
[95 % CI]	[38.2-67.5 %]	[39.2-59.3 %]	[46.3-54.6 %]
Primary School Education (true value = 5.3 %)			
Estimate	3.2 %	1.5 %	5.0 %
[95 % CI]	[1.6-6.4 %]	[0.7-3.1 %]	[4.1-6.2 %]
Without work (true value = 32.8 %)			
Estimate	51.3 %	35.4 %	36.8 %
[95 % CI]	[36.7-65.7 %]	[26.4-45.5 %]	[32.9-40.9 %]
Older than 65 (true value = 16.8 %)			
Estimate	17.2 %	14.1 %	15.6 %
[95 % CI]	[11.5-24.9 %]	[10.6-18.6 %]	[13.7-17.6 %]
Married (true value = 40.8 %)			
Estimate	33.4 %	44.5 %	41.0 %
[95 % CI]	[23.6-45.7 %]	[35.6-54.0 %]	[37.5-44.6 %]

NOTE.—Post-stratification weights applied.

It should be noted that both ABS samples are rather small and only contain 186 (address only) and 416 (augmented) respondents, respectively. This causes the confidence intervals to be rather large compared with those from the PBS sample, whose sample size is decidedly larger with 1,677 respondents. Therefore, coverage of the true population value by the confidence intervals is a tougher standard of judgment for the PBS sample.

We supplement our analysis and assess the magnitude of difference between true values and the population estimates. This further substantiates the results on the degree of nonresponse and coverage error. The estimates from the ABS sample have an average deviation from the population values of 6.2 percentage points. In comparison, estimates from the augmented ABS sample and the PBS sample are closer to the true values, with an average deviation of 2.8 and 1.2 percentage points, respectively. This means that although few point estimates came close to the true value, samples for which individual names were available performed better on average.

In summary, the two assessments lead to similar conclusions about which sampling frame designs are better in estimating the size of subgroups of the population. We can improve our confidence in the causal mechanism further when only considering the results obtained from the nonaugmented and augmented samples: the augmented sample derived from a sampling frame containing names and addresses of potential respondents performed better. These results are consistent with our argument that a sampling frame that includes

individual-level information provides better survey estimates because it reduces undercoverage.

5. CONCLUSION

This study systematically compared indicators of nonresponse and coverage error produced by different sampling frame designs commonly used for face-to-face interviews in the social sciences for a target population of individuals aged fifteen and over. We hypothesized that up-to-date availability of names in the sampling frame design could have as positive of consequences for the degree of nonresponse and coverage error as the kind of sampling frame design itself.

We conducted a unique quasi-experimental study in a single country to assess the differences between a person-based and an address-based sampling frame design. They are often considered the gold standard for sampling frame designs in theory and in practice, respectively. The results indicate that advanced availability of respondents' names could improve the nonresponse and coverage error. Even according to the most conservative measure, response rates, noncontact rates, cooperation rates, and ineligibility rates could improve by between 2.5 and 6 percentage points. However, it may be that the differences between the ABS address-only sample and the ABS augmented sample are confounded by selection biases. People whose names are unknown are also those who are more reluctant to be listed in other databases. Frequent movers, on the other hand, may be listed under an old address, which might explain differences between the PBS and the ABS augmented samples since the PBS frame is most likely to be up-to-date.

Our analyses also showed that person-specific information improves sample survey estimates of the size of specific subgroups in the population. A comparison of estimates for the composition of the Dutch population against true values revealed that estimates from the (augmented) person-based sampling frame design had an average deviation of between 1.2 and 2.8 percentage points, while the address-based design had an average deviation of 6.2 percentage points. This suggests that having person-specific information substantially improves sample survey estimates of specific subgroups. And even here, applying the most conservative standard of judgment still showed that when knowing respondents' names upfront, estimates and their confidence intervals covered the true population values in two out of three instances.

These results from this unique case study suggest that a PBS frame is not always necessary to minimize nonresponse and coverage error (see also Koch et al. 2014). According to our results, augmenting an ABS with names from another source already significantly improves several indicators of nonresponse and coverage error. This finding could be of considerable value for other researchers because a PBS frame is not always available, and its

acquisition might incur substantial costs on researchers and fieldwork companies. What is more, not all PBS frames, or frames in general, are of the same quality because of varying intervals for updating. In Germany, for instance, the most recent and first census since reunification took place in 2011, and the United States conducts a census every ten years. Additionally, register data collected at the municipal level may be less up-to-date if municipalities are not communicating with each other directly, for example, when individuals are moving houses. In such cases, a PBS frame could be less up-to-date than perhaps an ABS or HHBS frame based on recent electricity bills. This, in turn, might offset the positive effects of a PBS frame on the degree of nonresponse and coverage error. This means that the decision for a sampling frame design should also consider which frame and design provide the most updated and correct information about the target population in a specific country.

Our first results reported in this study need to be substantiated in further analyses and also in other countries. We acknowledge the limited replicability of our findings given that person-based sampling frame designs are not always available. This also highlights the potential value of our case study for future research. At the same time, in the ESS 2002–2012, out of the thirty-four countries that participated at least twice, eighteen had access to a person-based sampling frame design (see supplementary data online, table A.1). Future research could also further investigate the impact of employing different sampling frame designs through experimental studies (including a condition that randomly removes names instead of adding them). This may be a particularly difficult yet not impossible task. As this study showed, survey samples and the sampling process are exposed to many selection effects, making it difficult to disentangle effects and to establish truly experimental conditions for large-scale survey samples. Finally, future research could also use our matching technique of different samples to investigate the extent to which an address-based sampling frame design systematically captures an additional undercovered population.

Supplementary Materials

Supplementary materials are available online at academic.oup.com/jssam.

REFERENCES

AAPOR (2014), "Best Practices for Survey and Public Opinion Research," available at https://www.aapor.org/Standards-Ethics/Best-Practices.aspx#best3, last accessed: 07/25/2018.

DiGaetano, R. (2013), "Sample Frame and Related Sample Design Issues for Surveys of Physicians and Physician Practices," *Evaluation & the Health Professions*, 36, 296–329.

Eckman, S., and F. Kreuter (2011), "Confirmation Bias in Housing Unit Listing," *Public Opinion Quarterly*, 75, 139–150.

European Social Survey (2012), "Sampling for the European Social Survey Round VI: Principles and Requirements. Mannheim: Sampling Expert Panel of the ESS, GESIS," available at https://

www.europeansocialsurvey.org/docs/round6/methods/ESS6_sampling_guidelines.pdf, last accessed: 05/04/2017.

Farrell, D., and J. C. Petersen (2010), "The Growth of Internet Research Methods and the Reluctant Sociologist," *Sociological Inquiry*, 80, 114–125.

Groves, R. M., F. F. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2009), *Survey Methodology*, Hoboken, NJ: Wiley-Interscience.

Harter, R., M. P. Battaglia, T. D. Buskirk, D. A. Dillman, N. English, M. Fahimi, M. R. Frankel, et al. (2016), *Address-Based Sampling*, AAPOR. Oakbrook Terrace. Available at: https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx, last accessed 07/25/2018.

Iannacchione, V. G. (2011), "The Changing Role of Address-Based Sampling in Survey Research," *Public Opinion Quarterly*, 75, 556–575.

Kalton, G., J. Kali, and R. Sigman (2014), "Handling Frame Problems When Address-Based Sampling is Used for in-Person Household Surveys," *Journal of Survey Statistics and Methodology*, 2, 283–304.

Koch, A., V. Halbherr, I. A. L. Stoop, and J. W. S. Kappelhof (2014), "Assessing ESS Sample Quality by Using External and Internal Criteria," Mannheim, available at http://www.europeansocialsurvey.org/docs/round5/methods/ESS5_sample_composition_assessment.pdf, last accessed 05/04/2017.

Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn, and A. H. Mokdad (2006), "Address-Based Versus Random-Digit-Dial Surveys: Comparison of Key Health and Risk Indicators," *American Journal of Epidemiology*, 164, 1019–1025.

———. (2008), "A Comparison of Address-Based Sampling (ABS) versus Random-Digit Dialing (RDD) for General Population Surveys," *Public Opinion Quarterly*, 72, 6–27.

Link, M. W., and J. W. Lai (2011), "Cell-Phone-Only Households and Problems of Differential Nonresponse Using an Address-Based Sampling Design," *Public Opinion Quarterly*, 75, 613–635.

Särndal, C. E., B. Swensson, and J. Wretman (2003), *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Statistics Netherlands (2012), *Population*, available at https://opendata.cbs.nl/statline/#/CBS/en/dataset/37296eng/table? ts=1530263114934, last accessed 05/04/2017.

Stoop, I., H. Matsuo, A. Koch, and J. Billiet (2010), "Paradata in the European Social Survey: Studying Nonresponse and Adjusting for Bias," in *Joint Statistical Meeting 2010: Proceedings Section on Survey Research Methods*, pp. 407–421.

Tsuchiya, T., and N. E. Synodinos (2015), "Searching for Alternatives: Comparisons Between Two Sample Selection Methods in Japan," *International Journal of Public Opinion Research*, 27, 383–405.

Tourangeau, R., G. Shapiro, A. Kearney, and L. Ernst (1997), "Who Lives Here? Survey Undercoverage and Household Roster Questions," *Journal of Official Statistics*, 13, 1–18.

Tweede Kamer der Staten-Generaal (2009), "Modernisering Gemeentelijke Basisadministratie persoonsgegevens (GBA)," *Kamerstuk 27859 nr. 27*, available at https://zoek.officielebekendmakingen.nl/dossier/27859/kst-27859-27? resultIndex=52&sorttype=1&sortorder=4, last accessed: 07/25/2018.