



Catalogue no. 11-522-XIE

**Statistics Canada International Symposium
Series - Proceedings**

**Symposium 2004: Innovative
Methods for Surveying
Difficult-to-reach Populations**

2004



A PERMANENT SAMPLE AS A SAMPLING FRAME FOR DIFFICULT-TO-REACH POPULATIONS?

Thomas Körner and Anja Nimmergut¹

ABSTRACT

The trend towards a growing reluctance to participate in surveys challenges statistical institutes to find new approaches for sample construction. This is particularly true in the case of difficult-to-reach populations which against a background of eroding response rates could constitute a major obstacle for the nonresponse bias of general population surveys. The permanent sample of households ready to respond of the German statistical system constitutes one, albeit partial, possible solution to this problem. We refer to a permanent sample as a sampling frame composed of people ready to participate in voluntary surveys regularly. Certain types of surveys focussing on difficult-to-reach populations can be carried out more quickly and without a screening preceding the survey because up-to-date socio-economic information is available from the permanent sample. Furthermore, in many cases there should be fewer problems due to low response rates as the units in a permanent sample are generally ready to respond. This paper outlines some possible applications of the permanent sample of households ready to respond with respect to surveying difficult-to-reach population groups.

KEYWORDS: Access Panels; Household Surveys; Nonresponse Bias; Sampling Frames

1. INTRODUCTION

Nonresponse is one of the major challenges for household surveys. Problems related to the reluctance to participate in surveys of official statistics have become even more urgent in recent years. International research has made obvious that there is a trend towards decreasing response rates in many countries (see, e. g., de Heer, 1999; de Leeuw and de Heer, 2002; Schneekloth and Leven, 2003; Stoop, 2004a). Accordingly, most surveys have to find solutions of how to increase the response rates or at least of how to prevent a further decrease. However, it has to be noted that the overall response rate alone does not necessarily guarantee acceptable results. With respect to the nonresponse bias of a survey it is at the same time vital to motivate each of the different sub-populations of the target population to participate in the survey. Given the case that specific difficult-to-reach sub-populations are not well represented in the sample the nonresponse bias might well be high, while response rates seemingly have an acceptable level. In other words, the systematic loss of certain population groups should be regarded as a major hurdle for most voluntary household surveys.

Recent scientific meetings have shown that there is a general feeling that statistical institutes will more and more be forced to find new approaches for sample construction which go beyond the classical random sampling approaches available today. In response to the growing nonresponse problems and other recent challenges, the Federal Statistical Office Germany together with the Statistical Offices of the German Länder, in early 2004 has established a permanent sample of households ready to respond (*Dauerstichprobe befragungsbereiter Haushalte*). Although primarily introduced in order to obtain random samples of the general population at acceptable cost and to reduce the preparation time for ad hoc-surveys, the permanent sample at the same time can contribute to improvements of reaching some difficult-to-reach population groups (e. g. groups with low income, low formal education, single parents etc.). By accumulating households ready to respond over several recruitment waves, the permanent sample constitutes a possible way to oversample some difficult-to-reach populations or to draw samples for surveys focussing on (difficult-to-reach) sub-populations of the general population.

¹ Thomas Körner, Institute for Research and Development in Federal Statistics, Statistisches Bundesamt, 65180 Wiesbaden, Germany, thomas.koerner@destatis.de; Dr. Anja Nimmergut, Institute for Research and Development in Federal Statistics, Statistisches Bundesamt, 65180 Wiesbaden, Germany anja.nimmergut@destatis.de

This paper examines whether for some populations regarded as “difficult-to-reach” the permanent sample might serve as a sampling frame. First of all, we briefly introduce some key features of the permanent sample and outline its most important characteristics as a sampling frame. In the main part, based on the results of a pilot study, we examine which difficult-to-reach population groups are represented in the permanent sample. Finally, we show first response trends of these groups.

2. THE PERMANENT SAMPLE

The permanent sample of households ready to respond is a sampling frame for voluntary household surveys of official statistics. After a large scale pilot study, it has been implemented in early 2004 in order to counteract a number of challenges voluntary household surveys in German official statistics are facing, amongst others the growing reluctance of households to respond in surveys. The basic idea of the permanent sample is to permanently recruit households who have participated in the largest random sample of the general population in Germany, the microcensus. Those households who have been recruited successfully will be asked from time to time to participate in voluntary household surveys (for further details see Körner/Nimmergut, 2004).

Thus, the microcensus constitutes the basis for the establishment of the permanent sample and can be considered as the core of the system of household surveys in German official statistics. It is a multi-purpose random sample containing about 380,000 households being surveyed mandatorily every year. The microcensus provides statistical information on a large array of socio-economic variables (including comprehensive demographic and socio-economic information, the employment status as well as vocational and training issues). These are available in the context of the permanent sample and in turn can be used for effective weighting of the households. Due to its large sample size and a response rate of nearly 100 percent the results of the microcensus are considered highly reliable.

All households who have participated in their last microcensus interview (normally after four years) will be asked whether they are ready to participate from time to time in voluntary household surveys of official statistics. The permanent sample is introduced to the households under the product name “Households Today” (*Haushalte Heute*) as the term “permanent sample” would have been too technical to be used in the recruitment process. The microcensus interviewers address the households directly after the last microcensus interview.² The interviewers inform the households on the permanent sample and hand over an information leaflet explaining the participants’ role in the permanent sample in more detail. At the end of the interview, households have the opportunity to directly decide for a participation in the permanent sample, but they can also reconsider it and join the permanent sample later on declaring their consent via mail. All participants have to sign a declaration of consent which provides the legal basis for the storage of the information from the last microcensus interview into the data base of the permanent sample as well as to keep up-date some selected variables from this information every year.

In order to cover the sample of the microcensus entirely, printed information material is sent to the households who were not reached by the interviewer or who asked to complete the microcensus’ self-administered questionnaire. In this case, the printed information material is accompanied by a cover letter which introduces the request to the households from the head of the statistical office and asks the households to mail back the declaration of consent.

The permanent sample can serve as sampling frame for sample surveys of households and individuals which do not require a mandatory response and for which a comparatively small sample size is considered sufficient. Compared to other sampling frames for household surveys, the permanent sample is characterised by a number of special features which are essential for its effective use. As briefly mentioned, extensive socio-economic information is available from the last microcensus interview. This information can be used for an effective correction of the nonresponse bias occurring during the recruitment stage. In the context of the development of effective weighting techniques it is important to note that the socio-economic information from the microcensus is available for both participants and non-participants of the permanent sample.

² During the pilot study, this recruitment method has proven to be the one with the best cost-effectiveness ratio. In total, 18% of the households approached with that method participated in the access panel (Nimmergut/Meyer/Körner, 2004).

From the full set of microcensus variables, a subset of “core variables” is stored in the data base of the permanent sample and up-dated regularly. The core variables play an important role, as they can be used as stratification variables. In practice this means that for many surveys focusing on sub-populations of the general population it is not necessary to carry out a screening prior to the data collection. The core variables can equally be used in order to tailor the fieldwork to the requirements of different sub-populations. Finally, in the permanent sample, households participating are accumulated over the years. For this reason, the size of the permanent sample is constantly growing so that after few years also comparatively small sub-populations are represented with a sufficient number of households in the permanent sample.

3. DIFFICULT-TO-REACH POPULATIONS IN THE PERMANENT SAMPLE

In this paper we refer to difficult-to-reach-populations as groups of households and individuals who are systematically underrepresented in voluntary household surveys. The possible reasons for being “difficult-to-reach” are manifold, covering not only various socio-economic features but also, e. g., the geographical, legal, or health-related background of households and individuals. From the multitude of difficult-to-reach population groups covered by this definition only those can possibly be surveyed on the basis of the permanent sample for which a sufficiently large number can be contacted during the recruitment procedure. Especially very small populations, populations concentrated in only few regions, or populations which are difficult to define (e. g. native populations, immigrants, homeless people, or homosexuals) are generally rather unlikely to be sufficiently represented in the permanent sample. For difficult-to-reach populations which make up larger parts of the general population (like households with low income or low educational background, self-employed persons, or single parents), on the contrary, it is quite likely that the permanent sample can be used as a sampling frame.

To get an idea of which difficult-to-reach populations are well represented in the permanent sample one needs to compare the distribution of the respondents in the permanent sample with the respective distribution in the target population according to certain variables of interest. It has to be mentioned that from such a comparison evidently one can only identify such difficult-to-reach-populations which can be detected by the use of the variables available. However, socio-economic variables not in any case cover all possible “triggering mechanisms” for difficulties in reaching certain population groups. Recently, it is has been argued that even the most exhaustive socio-economic variables might miss “soft” factors which influence nonrespondents in their behaviour, such as interest in the survey, attitudinal factors (like the attitude towards official statistics) the general social and political climate (Harris-Kojetin/Tucker, 1999; Djerf, 2004) or even the respondents’ personality profile (Rammstedt et al. 2004). In the case of the permanent sample exhaustive socio-economic information is available for all microcensus participants (including the participants in the permanent sample), largely facilitating such a comparison.

Table 1 compares the distributions from the microcensus and the permanent sample pilot study for some population groups which are generally regarded as difficult-to-reach in voluntary household surveys (see also Körner and Nimmergut, 2004; Nimmergut and Körner 2004). The results show that the representation of the population groups in the permanent sample is varying according to the different socio-economic variables. Some groups, such as single parents or elderly women are even slightly overrepresented in the permanent sample. Other groups like self-employed persons, young men and households with low income, are slightly underrepresented in the permanent sample. Blue collar workers, and persons with low formal education are even more difficult to be recruited for the permanent sample.

Table 1: Representation of selected difficult-to-reach population groups in the microcensus and the permanent sample pilot study (Proportion of the population group in the total microcensus / permanent sample population)

Population group	Microcensus 2001	Participation of that group in the permanent samples' pilot study
Self employed persons	10,1 %	9,1 %
Blue collar workers	29,2 %	21,9 %
Age 20 to 29 (male)	5,7 %	3,5 %
Age older than 60 (female)	13,8 %	14,9 %
Single parents	3,4 %	3,9 %
Low household income (below 900€/per month)	13,2 %	9,6 %
Low formal education	51,4 %	41,6 %

4. INCREASING THE RESPONSE RATES OF THE DIFFICULT-TO-REACH POPULATIONS WITH THE PERMANENT SAMPLE

The most obvious advantage of the permanent sample are the higher response rates that can be obtained in the actual fieldwork of the surveys sampled out of the permanent sample: By recruiting households which are generally ready to respond in surveys, the response rates for samples drawn from the permanent sample (both for the general population as for difficult-to-reach population groups) tend to be significantly higher than in probability samples contacting the target population directly (Körner/Nimmergut 2004). This advantage can largely contribute to reduce the cost for the field work. With regard to the features of the permanent sample as a sampling frame, there are two further advantages which in the following will be presented in more detail: First of all the difficult-to-reach populations can be easily identified due to the availability of a variety of socio-economic variables. Secondly, due to the continuous recruitment process, an accumulation of households takes place which results in a pool of households where – although being underrepresented – a sufficient number of the difficult-to-reach households should be available, given that the sample is smaller than the overall permanent sample.

The easy identification of the difficult-to-reach populations in the permanent sample has two main implications. First of all, stratification sampling can be used very easy due to the availability of up to date socio-economic variables. They allow that the stratification approach can be adjusted flexibly to the requirements of each individual survey sampled from the permanent sample. Therefore, the precision of most surveys can be enhanced because in most cases stratification variables should be found that correlate with the estimates for the characteristics of interest. However, with respect to the difficult-to-reach populations stratification with oversampling is of particular interest. The permanent sample after several successive recruitment waves enables us to oversample difficult-to-reach populations in the gross sample to ensure that they are relatively well presented in the net sample, in other words in the final results. Of course, this approach is only promising if the net sample size of the difficult-to-reach populations' strata is substantially smaller than the number of households in the permanent sample. Given that this precondition is fulfilled the permanent sample has a big advantage concerning the difficult-to-reach populations due to the relatively easy possibility of stratifying with oversampling.

Another advantage is the general possibility to run surveys focusing on specific subgroups of the general population. This might be a particular advantage if one wants to learn more about specific difficult-to-reach-population groups. It does not directly improve the accuracy of the results of an individual household survey but could help learning more about the soft factors (see above) that may also contribute to be “difficult-to-reach”. In this case it enables the investigation and documentation of the behaviour of particular difficult-to-reach groups which would subsequently help to adjust fieldwork and methods for further surveys. One variable which at least might incorporate some attitudinal factors and which might help to detect specific subgroups with different attitudes is the occupational status. Its categories (e. g. officials, judges versus self-employed and blue collar workers versus employees) are a priori independent on the income, educational training and to a certain degree on the social status but exhibit nevertheless quite substantial differences in the response rates. This variable is already available from the microcensus and could be therefore easily used for comparisons with the total population.

Finally, the identification of the difficult-to-reach-populations prior to the actual fieldwork would allow organising the fieldwork in a way that aims particularly at reaching and motivating these specific groups. Beginning with the original recruitment from the microcensus the interviewers could be instructed to pay special attention to the difficult-to-reach populations, the interviewers could be trained in special strategies to approach them and of course specific follow up strategies could be employed, e. g. follow-up via telephone to create a second personal contact to blue-colour workers or people with lower educational background.

Presently, these strategies are only very scarcely employed due to organisational limitations and also the fact that the mixed mode methods itself might have biasing effects. However, if any such method proves to be efficient without having other negative effects on the survey results, the permanent sample is an instrument which could be regarded flexible enough to implement it relatively quick at relative low costs.

5. ADJUSTING FOR THE NON-RESPONSE OF THE DIFFICULT-TO-REACH POPULATIONS

The advantages for adjustment with respect to difficult-to-reach populations may not be that pronounced. The big advantage of course is the availability of socio-economic variables for the respondents as well for the non-respondents. This e. g. makes the estimation of individual participation probabilities for a participation in the permanent sample, remaining in the permanent sample and for participating in the voluntary surveys based on the permanent sample possible. The participation probabilities in turn could be used to construct weights which then would have to be combined for estimation. However, the details of the concept for estimation have not yet been fully decided and there are also problems associated with the design of the permanent sample with respect to adjustment. Accordingly the question if the permanent sample has advantages for the adjustment of difficult-to-reach populations has to remain open until the estimation and variance estimation concepts have been completed.

6. CONCLUSIONS

The results available from the pilot study suggest that the permanent sample could be very usefully applied for some surveys of particular difficult-to-reach populations. Its application is however restricted to population groups which have a sufficiently large size in order to be reasonably represented in the permanent sample. Such population groups include for example households with low income, persons with low formal education, elderly people, self-employed persons, or single parents. By the use the permanent sample the efforts necessary to reach such population groups can be reduced. However, using the permanent sample for surveying these groups is only possible if the required sample size is fairly small, and in any case distinctively smaller than the permanent sample as a whole. In many cases, the quality of the sample to be drawn will improve after accumulating households over several successive recruitment waves. Given that these preconditions are met, the permanent sample has a large potential for ad hoc surveys of such population groups or for oversampling these groups in general population surveys. With the possibility to carry out surveys focussing on sub-populations without a previous screening, preparation time and field cost could be considerably reduced. However, the permanent sample is not a useful tool for highly specific sub-populations, like ethnic minorities, homosexuals etc.

It has been shown that the strength of the permanent sample is the availability of extensive socio-economic information for both the participants and the non-participants. This advantage together with the strict access limitation to former microcensus households the permanent sample is a very strong instrument compared to many access pools used by commercial market and public opinion research institutes (see also Stoop, 2004b). With respect to difficult-to-reach populations the availability of socio-economic information is of outmost importance. First of all it enables an easy identification of the various population groups one might want to focus on. Secondly, it helps increasing response rates by tailoring fieldwork methods to the requirements of the different population groups. And thirdly, the available information enables us to develop effective weighting methods.

REFERENCES

- Djerf, K. (2004), "Nonresponse in Time: A Time Series Analysis of the Finnish Labour Force Survey", *Journal of Official Statistics*, 20, pp. 39-54.
- Harris-Kojetin, B. and Tucker, C. (1999), "Exploring the Relation of Economic and Political Conditions with Refusal Rates", *Journal of Official Statistics*, 15, pp. 167-184.
- de Heer, W. (1999), "International Response Trends. Results of an International Survey", *Journal of Official Statistics*, 15, pp. 129-142.
- Körner, T. and Nimmergut, A. (2004), "Using an Access Panel as a Sampling Frame for Voluntary Household Surveys. Results from a Pilot Study", *Statistical Journal of the United Nations Economic Commission for Europe*, 21. (in print)
- de Leeuw, E. and de Heer, W. (2002), "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison", in Groves, R. M. et al. (eds.) *Survey Nonresponse*, New York: Wiley, pp. 41-55.
- Nimmergut, A. and Körner, T. (2004), "Reducing Nonresponse, Enhancing Flexibility, Improving Quality in Household Surveys. The New Access Panel in the German Statistical System", *Proceedings, European Conference on Quality and Methodology in Official Statistics (Q2004)*, Mainz, Germany.
- Nimmergut, A. Meyer, I. and Körner, T. (2004), *Haushalte Heute. Pilotstudie zur Umsetzbarkeit einer Dauerstichprobe befragungsbereiter Haushalte in der amtlichen Statistik 2001-2003*, Wiesbaden: Statistisches Bundesamt.
- Rammstedt, B. et al. (2004), Entwicklung und Validierung einer Kurzskaala für die Messung der Big-Five-Persönlichkeitsdimensionen in Umfragen, *ZUMA-Nachrichten*, 55, pp. 5-28.
- Schneekloth, U. and Leven, I. (2003), Woran bemisst sich eine "gute" allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS, *ZUMA-Nachrichten*, 53, pp. 16-58.
- Stoop, I. (2004a), Surveying Nonrespondents, *Field Methods*, 16, pp. 23-54.
- Stoop, I. (2004b), "Access pools as a solution to the nonresponse problem?", paper presented at the *15th International Workshop on Household Survey Nonresponse*, Maastricht, The Netherlands.