Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

Data Documentation

33

Rainer Schnell

Avoiding Problems of Traditional Sampling Strategies for Household Surveys in Germany: Some New Suggestions

Berlin, June 2008

IMPRESSUM

DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1861-1532

Avoiding problems of traditional sampling strategies for household surveys in Germany: Some new suggestions

Rainer Schnell

Center for Quantitative Methods and Survey Research
University of Konstanz

31st March 2008

Abstract

All of the sampling plans currently in use for general population surveys in Germany suffer from methodological and practical problems. A new sampling plan is thus urgently needed: one with a low cost overhead that can be prepared in a very short time. Germany also lacks a sampling plan covering all institutional populations, immigrants in general, and illegal immigrants in particular.
The availability of new databases covering these populations suggests ways of developing, implementing, and testing new sampling plans for population surveys in Germany. One such sampling plan (G-Plan) is proposed here for the first time.[1] The implementation problems of this design must be studied in a number of empirical pretests.

1 Current sampling plans for national population surveys in Germany

Up to the present, sampling plans for demanding general population surveys have usually been carried out in face-to-face interviews. Mail surveys of the general population are virtually never used in Germany.[2] Internet surveys suffer from low population penetration and the complete lack of a sampling frame, and are thus never used for serious population research in Germany. CATI surveys are widely used, even for panel recruitment, but they have recently been shown to suffer from dramatically increasing nonresponse (due to telephone advertising). Of course, there is also undercoverage in CATI surveys due to non-telephone households (approx. 5%). Furthermore, at least some subpopulations cannot be contacted by landlines and there is no reliable research on sampling frames for mobiles phones in Germany. Therefore, sampling plans for CATI or Internet-based surveys will not be discussed here.

1.1 Problems of compiling frames and conducting fieldwork in Germany

The technical problems of sampling in Germany are usually underestimated by foreign statisticians. The main problems are:

- Germany does not assign individuals a national identification number (or social security number). This is in fact prohibited by the high court. Therefore, deduplication and linking of databases is difficult at best.
- There is no central registry of residents. Only municipal registries contain names and addresses.

[1] I am indebted to Mark Trappmann for a detailed critique of an earlier draft.

[2] The only exception seems to be a study in the DEFECT-project (Schnell/Kreuter 2000). Although the response rate was high by comparison, there is a clear educational bias in this survey.

- There is no household information in any official registry in Germany. Every registry contains only information on persons.
- Germany consists of 16 federal states. Each state has different laws on registries, independent data protection agencies, and statistical agencies. The federal statistical agency ("Statistisches Bundesamt") is not entitled to give orders to the statistical offices of the federal states.
- The municipal registries may or may not cooperate in sampling. They can decide independently according to their workload. The fees they charge for sampling vary widely.
- Depending on the kind of sample, more than a dozen data protection offices have to agree to the sampling procedure.
- Finally, there is no academic or federal field organization for national F2F-surveys. Fieldwork is done ad-hoc by federal statistical agencies for official statistics or by (currently two to four) commercial research companies.

1.2 Two traditional designs: random walk and municipal registries

The predominant sampling plans in postwar Germany, from 1945 to about 1971, were registry-based samples.[3] In 1972, a special design for the German Readership Sample was developed (the so called ADM design), which has been the blueprint for nearly all face-to-face-designs up to now.[4] In fact, only two variants of the basic version are in use today:

RW Random walk: selection of voting districts as primary sampling units (PSU), selection of households by random walk as secondary sampling unit (SSU), selection of persons within households by a Kish selection grid.[5]

EWA Selection of towns as PSU, selection of individuals from the municipal registries ("Einwohnermeldeamt").

1.2.1 Weaknesses of current designs

The traditional designs have been described in some detail (ADM (1979, 1999) and Arbeitsgemeinschaft ADM-Stichproben/Bureau Wendt (1994)), but a thorough methodological analysis is still lacking.[6] Furthermore, the necessary databases for drawing a

[3] The only exception are quota samples, which will not be considered further since there is only one institute in Germany that heavily advocates the use of these samples for general population surveys: The "Institut für Demoskopie" at Allensbach. It should be mentioned that these data are almost never used for academic research in Germany.

[4] A short history of sampling plans in Germany can be found in Schnell (1997) and Löffler (1999).

[5] The Kish selection grid is called "Schwedenschlüssel" in Germany.

[6] It is remarkable that the most detailed methodological analysis of the standard sampling procedure in Germany was done by a sociologist in 1982 in a widely ignored technical report (Hanefeld 1982),

sample of PSUs according to the traditional designs are not publicly available. The same applies to the sampling algorithm: the exact algorithm has never been published.[7] There is some evidence that even after sampling PSUs, the allocation of PSUs is manually "fine-tuned".[8] Therefore, neither analytical nor simulation studies on the estimated sampling error of the two designs have ever been published. In fact, the problem of inflated standard errors due to design effects have been widely ignored in the research in Germany.[9]

A surprising problem in German surveys is the definition of the population. For RW designs, the population is defined as the German population in private households. As Hanefeld (1982:8) notes, this is problematic in more than one way. This concept is not used[10]

- by official statistics in Germany
- for sampling PSUs
- for SSUs

Since the official statistics in Germany nearly always use persons in private households, published figures from the official statistics can rarely be used for sampling or weighting.[11] The population numbers based on official statistics must be computed from the

of which only six copies can be located in German libraries today. The study by Hanefeld was prepared for the survey that finally became the GSOEP.

[7]For example, in the RW design, the PSUs are voting districts ("Stimmbezirke" or "Wahlbezirke", but not "Wahlkreis"). Since they vary in size, small districts are merged; reducing their number from approx. 80,000 to 62,000. The merge algorithm is unknown.

[8]Although never documented by the German research companies, the actual selection of PSUs in both designs is not always a random process: The author has witnessed the replacement of some selected PSUs by others if the resulting sample of PSUs appears to show regularities, for example, local clustering. With the usually small number of PSUs (120-160), this may have some justification. Finally, sometimes PSUs are replaced by "similar" PSUs if the research company has no interviewers in a PSU.

[9]The first attempt to separate design effects and interviewer effects in German surveys was the DEFECT project (Schnell/Kreuter 2000, 2005).

[10]Hanefeld also notes that the concept is not used for selection of persons within households ("Sie wird bei der Stichprobenziehung der zweiten Stufe, bei der Auswahl der Haushalte und dabei bei der Bestimmung der Haushaltsgröße, nicht eingehalten").

[11]The details of household definitions and especially the small differences between persons at their main place of residence ("Ort des Hauptwohnsitzes") and persons at the main residence of the household's main income earner (" Ort des Hauptwohnsitzes des Haupteinkommensbeziehers") are quite tricky. The main difference is the counting rule for persons with more than one residence. Even if the definitions may be clear, the implementation in surveys is not: persons with more than one residence nearly always get the same weight as persons with only one residence. Therefore, population size may differ by approximately one million persons due only to different counting rules. I have to thank Bernhard Schimpl-Neimanns for many discussions of details in German official statistics. His website http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_2005/Definitionen/mz05_def_index.htm is an invaluable resource for definitions in German official statistics.

scientific use files of the micro census.[12]

The RW design uses voting districts as PSUs. The PSUs are selected by a systematic sampling procedure, thereby implementing a PPS design. The size of the PSUs is given by the number of households within them (Behrens/Löffler 1999:76). But since the number of households per PSU or town is unknown in German official statistics, the number must be estimated for each PSU (Hanefeld 1982:21), normally based on the micro-census. This introduces a possible bias in selecting "atypical" PSUs concerning the ratio of households to voters. Furthermore, it will increase the sampling variance.

1.2.2 Common problems of EWA and RW designs

Both traditional sampling plans have serious methodological problems in practice. The most important problem is the undercoverage of both designs.

Foreigners More than 6.7 million foreigners live in Germany. This is 8.1% of the population. Interestingly, this population is usually either excluded by design (as in registry samples) or partially excluded by interviewers, who report perceived problems of communicating with foreign respondents in German. In very few studies are native-speaker interpreters or translated questionnaires available to help foreign respondents from the main immigrant groups. In fact, there is no standard systematic approach. And since 8 million Germans (nearly 10% of the population) are recent migrants ("Deutsche mit Migrationshintergrund"), it is becoming increasingly difficult to interview even German citizens due to language problems. Since this population cannot be identified using data available in municipal registries, they can only be identified through interviews or screening.[13]

Institutional populations According to the Federal Statistical Office, 1% of the population (856,000) belongs to the institutional population (Statistisches Bundesamt 2006:11).[14] This population is extremely heterogeneous. We have (for example): nursing homes and residential homes for the elderly (more than 11,000), residence halls for students (1100), youth centers, shelters for battered women (350), residential homes for mothers and

[12] One highly disturbing fact in German official statistics is the weighting procedure of the micro census. The sample should be a cluster sample of 1% of the population. Since the sample is stratified, clustered, post-stratified, and reweighted according to external data (GREG), the weights in the scientific use file show remarkable variation. The validity of the external data used by GREG is not without doubt. For some details, see Heidenreich (1994), Afentakis/Bihler (2005) and Iversen (2007).

[13] Although Germany has a central registry for foreigners, due to legal constraints, this register cannot be used for sampling purposes.

[14] 3.5% of them have a regular dwelling outside the institution (30,000).

their children, residential homes for migrants, apprentice hostels, hospitals, hospices (20), nurses' hostels, homes for the blind, cloisters (3000), prisons (199) and military barracks (400). Due the distinctive features of German official statistics, the exact number of institutions and the number of persons living in them is not known.[15]
Due to practical selection problems, the indirect and difficult access to people in institutions, and the chance that it will be impossible to carry out an interview at all, this population is nearly always disregarded in general population surveys. For social research on the general population, this practice is unacceptable, especially if statements on life histories are needed. Therefore, if the institutional population is included at all, this is usually done in panel studies where the designated respondent is interviewed after he has left the institutional population.[16] Very few special purpose surveys include the institutional population by restricting the sampling universe to one special type of institution. Only the German micro-census includes the whole institutional population by the definition of a special stratum ("Gemeinschaftsunterkünfte").[17]

1.2.3 Special problems of random walk samples

Within the RW design, the selection of households within a PSU is based on a random walk of interviewers. In practice, this random walk is a doorbell panel sample. The main problem with the random walk version of the ADM design is the lack of interviewer control. The very few available studies on the actual performance of random walks indicate a low amount of repeatability and the avoidance of all non-standard quarters by the interviewers (Schnell 1991). Furthermore, the selection within households seems to be heavily biased (Sodeur 1997, Kohler 2007). Finally, in practice, field substitution of hard-to-contact respondents has been observed by simply extending the random walk; this error is even marketed by one German research company as a "random plus".

1.2.4 Special problems of registry samples

Due to problems of the random walk, the use of the registry-based samples is currently considered as the gold standard in German surveys. But recent tests of the German Census (Statistische Ämter des Bundes und der Länder 2004) demonstrate that the

[15] The only attempt to quantify this population was done by the present author (Schnell 1991). The numbers above are estimates based on my own recent research. In most cases, the numbers are lower limits. In order to validate the estimates, I counted the occurrence of different keywords for each type of institution in a database of all German phone numbers. The resulting lists were de-duplicated and individually checked by a research assistant.

[16] This rarely happens in practice, however, unless the respondent returns to a panel household.

[17] The technical definition of this stratum is based on the 1987 German census and continuous municipal reports on building activity. This will result in an underestimation of the institutional population, especially of small institutions. The situation will improve after the German census of 2011, but due to the proposed methods for the German census, even then the undercounting of special populations like illegal migrants will persist.

registries suffer from over-coverage and under-coverage; the amount of these errors increases with population size of the towns. In large towns, the overcoverage exceeds 6%; the maximum undercoverage is even higher. The methods used to detect the coverage problems are far from perfect, and thus the actual coverage problems are likely to be larger. The main problem of registry-based samples are foreigners, especially if they do not work as employees.[18] In practice, the undercoverage of town registries will yield too few students in university towns and a near complete missings of illegal migrants.[19] Therefore, at least in large towns, the registries can not be considered complete. The overcoverage will contain many old people (already dead) and many very young people (probably living with their parents).

The selection of registry-based samples is quite costly and time-consuming because every selected town must be contacted, negotiated, and paid separately.[20] Actually carrying out such studies is quite frustrating: in practice, some towns are never sampled, while others are replaced due to the time constraints of the studies. There has been no study of the possible "nonresponse" or substitution effects of the PSUs.

2 The sampling plan for the IAB household panel (PASS)

Since the existing sampling plans suffer from severe problems, new sampling plans that can be used with low overhead and fewer coverage problems and biases have been sought for some time by ADM and research agencies. Up to 2006, no new sampling plan has been proposed.[21] In 2006, the German Institute for Employment Research (IAB) asked for a review of possible sampling designs for a low-income household panel in Germany (Schnell 2006a). It became evident that none of the suggested sampling plans could be implemented without extensive preparation, making the survey useless for the intended purpose of monitoring effects of recent changes in the German social security system. I therefore suggested a new dual-frame sampling plan (Schnell 2006b), where a subset of the population within a PSU is sampled with probability proportional to their estimated ability to pay for credit (as reported by a commercial agency). This sampling plan has been accepted by the IAB for their household panel (n=26,000). The fieldwork of the

[18] A recent comparison of the town registries with the central registry for foreigners (AZR) showed a difference of 534,000 persons (Opfermann 2006).

[19] Estimating the number of illegal migrants is nearly impossible with existing databases in Germany. Nevertheless, the literature gives numbers between 100,000 and one million. Recent estimates report higher numbers within that range, with 500,000 as the lower limit. For details, see Lederer (2004) and Sinn et al. (2005).

[20] The selection of a proper registry-based sample for the federal republic may take more than six months. Depending on the number of PSUs, the cost for the sample alone may be between 40-60,000 euros. Details of the actual problems have been published by Albers (1997).

[21] The only new idea for these kinds of samples – and one that appears never to have been used in practice – is to conduct a random walk based on telephone directories (Marhenke 1997).

first wave has just been completed.[22]

3 New sampling frames and a new sampling plan for general population surveys

In a grant application submitted to the German Research Foundation (DFG) in August 2007, the author proposed a new sampling plan for national surveys of the general population (G-Plan). Neither foreigners nor institutional populations have to be excluded, and the plan can be adopted easily if desired. Furthermore, it avoids many of the problems of the traditional design.

3.1 A new frame: buildings

Since April 2006, there exists a new official database of all buildings in Germany. This database will permit selection of buildings with a clearly defined selection probability. A sampling plan based on such a frame has never before been proposed for a German survey.
Currently, two geo-referenced databases of all buildings in Germany are available: one from an engineering company (www.killetsoft.de) and the other from the 16 official German mapping agencies (www.lverma.nrw.de). The official database with 18.1 million entries will be available for academic use at 105,000 euros, the commercial database with 19.9 million entries at 28,000 Euro.[23]

3.2 Selection of PSUs: towns

The simplest way to sample from either frame would be an SRS of houses. Depending on survey requirements and cost restrictions, stratification and clustering can be done at any desired level. The details of this step should be explored in advance of actual sampling by simulation. Almost certainly, a stratification according to federal states may be useful or – depending on the purpose of the study – even necessary. In order to keep the travel costs for interviewers down, clustering is unavoidable for F2F surveys. Therefore, the selection of towns as PSUs is recommended. Since the number of households per building vary between towns of different population sizes, a stratification according to

[22] Due to time and money restrictions, the IAB decided that the panel would not be weighted according to the original proposal, so comparisons of the results of this panel with other samples would be misleading. Further details on PASS can be found in Promberger (2007).

[23] Since the commercial database claims independent measurements, it may also include shelters, garages, and other minor buildings. Use of this database might require an independent verification of the building type. The commercial database was derived by interpolation of GPS mapping data, so the given geo-code refers to a point near a building, not to the building itself as in the official database. Use of this database may therefore require additional decisions in the on-site selection of households.

population size of PSUs may be useful. Finally, a PPS sample seems to be a far better choice than a SRSWOR within each stratum.

3.3 Selection of SSUs: buildings

The main problem for sampling in Germany is the estimation of sampling probabilities within PSUs: since laws, database formats, database types, retrieval systems, and aggregation levels vary between federal states and municipalities, the population number of every aggregate unit below the town level is difficult to obtain in practice.[24] It is therefore suggested that this step be omitted entirely. Within each PSU a fixed number of buildings is selected SRSWOR from the sampling frame.[25]

3.4 Selection of households

Currently, the number of residential buildings is estimated by the German Federal Statistical Office at 17,458,815, with 38,587,180 dwellings. 62% contain one dwelling, 20% two dwellings and 18% three or more dwellings. Therefore, in more than 60% of the residential buildings, the building contains only one dwelling. In these cases no selection within buildings is needed.

All other buildings require the selection of households. The selection of one household per building is suggested.[26] I suggest the following procedure: the doorbell panels can be photographed by a simple mobile phone.[27] The photo can be sent per MMS to the supervisor, who does the selection by SRSWOR and sends the name (or the number on the doorbell panel) to the interviewer by SMS. Since the photographing can be disguised as the use of a mobile for writing an SMS, the photographing can be done unobtrusively.[28] This selection of households avoids any biasing liberal interpretations of selection rules of households. This is especially true if the household selection is done with different staff in advance of actual interviewing.[29]

[24] I have to thank Barbara Erbslöh (Office of Statistics and Elections, Essen) for the discussion of details of German municipality registers.

[25] In order to include buildings that are not in the frame, the technique of half-open intervals is recommended: new buildings between the selected building and the next not-selected building are included. These new buildings will get the same selection probability as the selected unit.

[26] I have no information on the distribution of the number of dwellings within a building if the number exceeds 10. In order to prevent excessive weights, it might be necessary to split buildings with a larger number of dwellings into smaller units.

[27] This idea was proposed by Menno Smid (Infas, Bonn). Some experiments have shown that a 2M-pixel camera yields photos that allow identification of names on doorbell panels in nearly 80% of all residential households. On the remaining panels, the layout of the panel allows a simple random sampling procedure.

[28] By comparison, even counting names on doorbell panels has caused problems with residents in previous projects by the author.

[29] In order to minimize interviewer cheating, the interviewers might be ordered to take photographs of the buildings, which can be compared with photographs made independently.

3.5 Selection within households

Currently, we have an estimated number of 39,178 million households in Germany. 14.7 million households consist of only one adult (37.5%), 12.8 million households contain two adults (34.1%). The selection within single households is unnecessary and therefore without error. It is suggested here that the interviewer ask for the number of adults within households. The interviewer enters this number in a CAPI system. If the number of adults is two, the system will select at random with equal probability either the contact person or the other adult. In the remaining 28.4% of households, the interviewer collects a household roster of all adults, with a standard sequence (for example, oldest male to youngest female). The CAPI system should not allow any corrections at this stage of the interview in order to prevent any tampering during selection.[30]

3.6 Estimated inclusion probabilities of persons

The multi-stage sampling mechanism above can be used for the estimation of inclusion probabilities p_i of persons by using the inclusion probabilities of buildings p_b:

$$p_b = p_{PSU} \times \frac{number\ of\ selected\ buildings}{number\ of\ buildings\ in\ PSU} \tag{1}$$

$$p_i = p_b \times \frac{1}{n\ of\ HHwithin\ building} \times \frac{1}{n\ of\ persons\ within\ HH} \tag{2}$$

The p_{PSU} can be based on the number of persons per PSU (this requires at least one number based on official statistics) or on the number of buildings within a PSU. In the latter case, no information from official statistics has to be used, which makes the design attractive for validation studies.

Unfortunately, a few problems remain. Since the sample is actually a doorbell-panel sample, the problems of multiple residences remain. It should be tested empirically by using additional questions in the survey if weighting according to reported multiple residences will improve estimates. It must be stated that this weighting will suffer from the missing information of nonresponding households. Finally, the estimated number of household members relies on the (unreliable) data reported by respondents. It should be noted that an RW design suffers from the same problem.

3.7 Selection within institutions

If a survey of the general population is intended, we can simply use the buildings frame. If a building is classified as an institution (either by comparing to a list or by local

[30] Since tampering during selection is a common and hard-to-detect variation of interviewer cheating (Schnell 1991b), the selection must be monitored and validated by the usual interview validation. It is of utmost importance that selection errors of interviewers are sanctioned with high probability. In contrast, even the ALLBUS does contain known cases of selection error. Interviewer sanctions due to selection of wrong respondents are not mentioned in the survey reports on the ALLBUS.

inspection), we have to act according to the type of institution. For example, residence halls for students can be handled as residential buildings. Perhaps we have to exclude some institutions like military barracks.[31] Which institutions can be handled by which procedure must be empirically checked by a pilot study of the suggested designs.

If the focus of an intended survey is the institutional population, a different way of sampling seems necessary. In order to sample from the institutional population, two different approaches appear possible to select the PSUs:

- random sampling from the building frame, screening by local inspection
- compilation of a list of institutions, stratification by type, random selection within strata

Since we assume between 1 and 2 % of the population within institutions, the number of persons in a standard-size survey would be at maximum about 60 if we use the first approach. This is insufficient for parameter estimates of subpopulations within the institutional population. If such parameter estimates are needed, we have to use the second option.[32] In each case, more than one person per PSU should be selected by a random procedure. It should be clarified for each institution identified if a random sample within the institution is possible. If this cannot be done, I suggest a systematic name-based sampling procedure.[33]

3.8 Inclusion of foreigners

Surveys face three types of problems with foreigners:

- they are a rare population,
- the sampling frames are incomplete,
- they may require special interviewers or questionnaires.

If we want to include foreigners, we must solve all three problems. The usual solutions (screening, snowball samples) confound these problems.

[31] Even in this case, we can use the number of registered persons in the municipal registry and use this number for weighting. Furthermore, for a panel we can consider a yearly check if the person moved out, and track them outside. The legality of this procedure must be checked for different institutions and all 16 different laws of registration within the Federal Republic.

[32] Since we must test the operational procedures for institutions in a pilot study, we need a large number of institutions in the pilot. Therefore, a complete list of institutions should be compiled. Using publicly available information, such a preliminary list can be compiled within a week (this has already been done by the author's workgroup). From this list, a large sample of institutions of different types should be selected.

[33] Such a scheme is also used in the German microcensus. The groups for the sampling are still based on the 1987 German census. An updated grouping has been prepared by the author based on all names in the databases of the German unemployment agency.

3.8.1 Solving the frame problems for a rare population

If one intends to collect data on foreigners, either screening procedures (Kalton 2001) or registers are used.
Screening is very expensive and may be biased (Granato 1999). Registers are hopelessly incomplete (for example, due to illegal migrants) and usually outdated by a few months. In order to solve the problem of rare occurrences and incomplete frames, I suggest the application of a linked sampling design (Levallee 2007):
If a selected household turns out to contain at least one foreigner, [34] it will be included. If a high number of foreigners is required, it may be possible to use the classification probabilities for a disproportional sample.

The number of foreigners in the sample might be increased further by two different techniques:

best friends All adult foreigners within a selected household could be interviewed. By asking each adult for his best friend of the same ethnic background in town, one can try to include them (without continuing the sequence to prevent a snowball sample). If there is no best friend in town of the same ethnic background, this information must be used for computing weights.[35]

neighbor If a selected household does not contain at least one foreigner, we could ask in buildings with exactly two dwellings whether the other household contains a foreigner. In this case, we have a defined link (Lavallee 2007) and can estimate the selection probability. The same technique can be used in buildings with more dwellings; in these cases we will ask for the nearest neighbor with a standard sequence: for example, opposite door, left, right, above, below.

So we have three different ways of obtaining a link to a foreigner, one direct and two indirect. At least the neighbor-based indirect links seem to be of acceptable reliability, so weights can be computed. Of course, this must be empirically checked by a pilot study.

3.8.2 Translation of questionnaires and special interviewers

The G-Plan requires an initial visit by an interviewer or surveyor to each selected building. This visit can be used – at least for some ethnic minorities– to make appointments

[34] The author's workgroup has developed a Bayes classifier for nationalities, employing the names used with the record-linkage software (MTB) under a DFG-funded project (Schnell/Bachteler/Reiher 2005). The conditional probabilities are based on a German database with a few million entries. The performance of this classifier based on doorbell panel entries can be compared with interview data during the pilot.

[35] It is unknown whether the "best friend" link is reliable and stable over time (at least during the survey period). Furthermore, we do not know if we can omit the problem of multiple links to the same person: we therefore have to collect all names (and birthdays, if possible) and count multiple occurrences.

with specialized interviewers. Alternatively, the interviewer selects the household, determines the required language, calls the institute and gives the mobile phone to the respondent, who is interviewed on the phone by a native speaker of his language. I have not found this procedure in the literature, and an empirical check will be needed in a pilot study to evaluate the feasibility of this procedure. But at least it combines the advantages of both interview procedures at an acceptable price.

4 Elements of a pilot study for the sampling plan

In order to study the feasibility of the suggested sampling plan, some empirical studies including a pilot implementation of the design are necessary.

4.1 Over- and undercoverage of different databases for sampling frames

At first, the databases available for constructing sampling frames must be empirically compared. The use of multiple additional databases (for example: telephone directories and direct marketing databases) in combination with on-site inspections will permit the selection of households and persons within households by the use of multiple-frame selection techniques (Coburn/Warde 1980, Deming/Glasser 1959). Multiple-frame plans have never been used or proposed before in German F2F surveys. The resulting selection technique could be compared and adjusted by the use of official registry data and independent enumerations or estimates within selected buildings.

4.2 Specification of the basic design

Although the database permits an SRS of houses, it should be studied in detail whether clustering (for example by communities), stratification (for example by federal state and/or community size) or PPS instead of SRSWOR of PSUs is more efficient. This should be done by simulating different design options. Furthermore, even after fixing the other sampling stages, the optimal local clustering within PSUs must also be examined by simulation: in order to keep design effects small, clustering of addresses should be avoided; in order to keep costs down, a certain amount of clustering is desirable.[36] Finally, the design weights for including foreigners and institutional populations must be described and computed in detail.

[36] Given a sampling frame of geo-coded buildings, clustering can be done by randomly selecting initial buildings, clustering with a standard algorithm according to euclidean distance, and random sampling within chosen thresholds. The implementation within a program like STATA is trivial.

4.3 Sample selection

At least 160 PSUs should be selected (probably by PPS) (frame: list of municipalities). Within each PSU, 5 streets will be selected (by SRSWOR; frame: building registry). Assuming a response rate of 40%, we will need 40 addresses per PSU (selected by SRSWOR from the building registry frame) to obtain 16 respondents.[37] We will take photos of doorbell panels. Using different databases as a replacement for independent multiple enumerations, the standard techniques for the completeness of census enumerations can be applied ("capture-recapture" models). These techniques have never been applied for the construction of sampling frames in Germany. At each selected address, households and persons within households will be sampled as described in detail above.

4.4 Evaluation of different building sampling frames

If 160 PSUs and 5 streets are selected within each PSU, the suppliers of the building databases can be asked for complete listings of 30 streets. Assuming only 5 buildings per street, at least 24,000 buildings can be compared. This will give a power of > 0.9 to compare two proportions (0.01 vs 0.015) of missing or wrong coordinates. Each list will be compared with address lists from other databases, for example, the phonebook and the IAB-registry data. Furthermore, the building coordinates for some of the selected PSUs can be compared by comparing the geo-codes with high-resolution pictures on the Internet (Google Earth, Visual Earth).[38] Finally, an onsite inspection of the selected 5 streets at 160 points is necessary. The enumeration of buildings and the classification of building types will allow a detailed examination of the database coverage.

4.5 Feasibility study of the procedure for sampling households

For a selected address, the enumeration will result in a photo of the doorbell panel. The number of households at this address can be compared with independent databases: phone directories, address resellers, and town registries.[39] Finally, the respondents (or in case of nonresponse: neighbors) should be asked for the number of households in the building and adults within the selected household.

4.6 Feasibility study of the procedure for sampling persons

Selection within households according to the CAPI selection grid should be without error if the number of adults is specified correctly. The number of adults within a household

[37] 160 PSUs with 16 respondents gives us n=2560; a common sample size in Germany (for a computation of power, see below).

[38] In those rare cases in which a high-resolution aerial photo of a selected street is not available on the Internet or at a local government office, it might be possible to obtain additional photos from commercial aerial services. Photos for a cluster of streets can be obtained for 350 euros per cluster.

[39] Town registries contain only persons, not households. For validation, it would be advantageous to have the number of persons and if possible, their names and the date of moving in: these variables are used by German official statistics to generate households.

should be obtained from the town registries.[40] Unfortunately, differences may arise due to interviewer cheating or respondent error. I suggest recontacting the respondents by mail and asking for the number of adults within a household. Nonrespondents to this question that can be found in phone directories should be contacted by phone.

4.7 Compilation of a list of institutions

A complete list of institutions can be created by using lists of umbrella organizations (for example, "Studentenwerke" or student organizations), special address books (for example, "Altenheimadressbuch" or retirement home address books), business directories ("Yellow Pages") and commercial databases ("Amadeus"). Furthermore, social security system data could be used to identify unique work places according to their economic sector.[41] The main problem here is the de-duplication of the list. Since de-duplication is a special case of record linkage, in which the author's workgroup has specialized experience, I do not expect any problems here.[42] The quality of the resulting list should be evaluated by comparing every building within a street of a selected building with the list. Furthermore, every administration of an institution within a street of a selected building should be contacted and asked for the number of residents. During this interview, asking for similar institutions within the same town known to the administration will yield further institutions. The resulting de-duplicated and verified additional list can be used in capture-recapture calculations in order to estimate undercoverage and overcoverage of the list.

4.8 Feasibility of sampling within institutions

Every administration of an institution within a street of a selected building should be contacted and asked for the number of residents. If no institution can be found within the street, the geographically nearest institution should be selected. The given number of residents should be compared with the number in the town registry (currently, we do not know anything about the feasibility, costs, or accuracy of this procedure). It should be tested in every selected institution if a SRS of residents within the institution or a systematic name-based selection procedure is feasible.

4.9 Implementation of a survey using the new plan

In order to evaluate the suggested sampling procedure, an empirical study seems to be necessary. It is suggested to conduct a survey with a sample selected by the G-Plan.

[40] This may be possible in some federal states and not in others. These problems must be discussed with the data protection agencies of the federal states and each town.

[41] The Research Data Center of the IAB has agreed to cooperate with the author in building this list.

[42] The development and implementation of record-linkage algorithms is subject of a series of projects I have directed since 2003. The resulting software ("MTB") is now being used, for example by Institute for Employment Research (IAB) and cancer registries in Germany, the Ukraine, and the UK.

A large part of the survey questionnaire should consist of factual questions, and those distributions can be compared to marginal distributions of official statistics or small area aggregate statistics available from the German unemployment insurance agency. The remaining parts of the questionnaire should consist of questions adopted from high-quality surveys conducted at the same time. Very few surveys in Germany demand (and pay for) the highest possible quality of data collection. Examples of such surveys include GSOEP, the CAPI sub-sample of PASS, the "Media-Analyse" (the CAPI sub-sample), ESS, and ALLBUS. After adjusting for obvious differences in the target populations, marginal distributions of the new survey with these high-quality surveys can be compared with the new sample. Of course, multivariate models should be also be compared, for a example, by a multi-group SEM. If we want to achieve a 95%-confidence interval of a binary variable with a 50% proportion, a resulting CI of 10% and an assumed design effect of 1.4[43], we need 2560 observations. Using a usual number (for Germany) of 160 PSUs, we will have 16 observations per PSU. Assuming a response rate of 40%, we need to sample 40 addresses per point or 6400 addresses in total.

4.10 Costs of the study

A professionally conducted F2F-survey (CAPI, 4 callbacks, contact protocols, 30 minutes, maximum workload 12 interviews) will cost about 80-100 euros per case plus overhead. Assuming 100 euros per case and 16 cases per PSU, the actual survey costs will be 256,000 euros at least.

A subsample of the two geo-databases of buildings in Germany is needed. This will require 15,000 euros in total. Furthermore, additional special databases, like the yellow pages, business address books, and so on are needed. 2000 euros seems sensible for these databases. Finally, it is highly recommended that whatever information for selected streets can be obtained from commercial companies like "Microm" or "AZ-Direkt" should be bought. They supply information on foreigners, inhabitants, and even names. Depending on the number of addresses within selected PSUs, the price for each database will be around 8,000 euros. In order to compare the geo-codes in the database with the actual coordinates, for each surveyor a GPS that can be linked to laptop and a camera is needed. In order to evaluate 100 buildings in 160 PSUs during a month, 8 units are needed. Suitable units (for example, Garmin Geko 201), the necessary cables and the software for geo-coding photos will add up to 2000 euros for GPS equipment.
Finally, the services of the municipal registries have to be calculated. As already mentioned, the price depends on the specific prices made by selected towns. Is seems to be safe to assume 400 euros per PSU or 64,000 euros in total. Therefore, the estimated total for databases and equipment is 99,000 euros.
Finally, every selected PSU must be visited for on-site inspection.[44] The surveyor

[43] This is the mean design effect in the first comparative study of design effects in German surveys, see Schnell/Kreuter (2004).

[44] The author has carried out a very similar procedure within his DEFECT project. All estimates are

must classify each building within a selected street as institutional, residential, or non-residential building. Furthermore, he should photograph each building with a geo-code so we can check the frames for completeness. Additionally, the surveyor should photograph each doorbell panel for the selection of households. Finally, the surveyor should obtain the number of residents within an institution and perform a random selection of households. A surveyor will need a day per PSU for his work.

A surveyor should work 6 days per week, 2 weeks in total. So we have 12 days and 12 PSUs per surveyor, and will need 160/12 = 13.3 => 14 surveyors. Each surveyor must be paid for work, travel, and accommodations for 12 days. Assuming 60 euros accommodation and 50 euros travel per day, we have to cover 110 euros per PSU: 17,600 euros. Paying each surveyor 1000 euros (12 days, 10 hours, 8.33 euros/h) adds up to 14,000 euros for surveyors. Therefore, the on-site inspection will cost 31,600 euros at least.

In order to prepare the data files, perform the linkages, check linked and non-linked cases in order to prevent artifacts by selective linkage, describe and analyze the data, highly qualified data analysts are required. In order to process the large number of databases, supervise the on-site inspections, negotiate with administrations, analyze the results, and compute weights, at least two Ph.D. candidates and four student research assistants are required. One Ph.D. candidate should work on the practical problems of field work, the other on simulating, modeling, computing weights and point-estimates, and comparing different databases. One student research assistant should be allocated to the latter, the other 3 to the fieldwork part of the project. Assuming a duration of 18 months, 160,000 euros for scientific staff is required.

This will add up to nearly 550,000 euros in total.

References

ADM (Arbeitskreis Deutscher Marktforschungsinstitute) (1979): Muster-Stichproben-Pläne, München.
Arbeitsgemeinschaft ADM-Stichproben/Bureau Wendt (1994): Das ADM-Stichproben-System. Stand: 1993; in: Gabler,S./Hoffmeyer-Zlotnik,J.H.P./Krebs, D. (eds.): Gewichtung in der Umfragepraxis, Opladen, 188–202
ADM (1999): Stichproben-Verfahren in der Umfrageforschung, Opladen
Afentakis,A./Bihler,W. (2005): Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005; in: Wirtschaft und Statistik, 10, 1039-1048
Albers, I. (1997): Einwohnermelderegister-Stichproben in der Praxis. Ein Erfahrungsbericht; in: Gabler, S./Hoffmeyer-Zlotnik, J.H.P. (ed.) Stichproben in der Umfragepraxis, Opladen: Westdeutscher Verlag, 117-126

based on the resulting experience.

Behrens,K./Löffler,U. (1999): Aufbau des ADM-Stichproben-Systems; in: ADM (1999): Stichproben-Verfahren in der Umfrageforschung, Opladen, 69-91
Coburn, T. C./Warde, W. D. (1980): A Sample Survey Design to Estimate the Means of Two Overlapping Subpopulations; in: ASA Proceedings of the Section on Survey Research Methods, 188-193
Deming, W. E./Glasser, G. J. (1959): On the Problem of Matching Lists by Samples; in: Journal of the American Statistical Association, 54, 286, 403-415
Granato, N. (1999): Die Befragung von Arbeitsmigranten: Einwohnermeldeamt-Stichprobe und telefonische Erhebung? ZUMA-Nachrichten, 45, 44-60
Heidenreich, H.J. (1994): Hochrechnung des Mikrozensus ab 1990; in: Gabler, S./Hoffmeyer-Zlotnik,J.H.P./Krebs, D. (eds.): Gewichtung in der Umfragepraxis, Opladen, 112-123
Iversen, K. (2007): Auswirkungen der neuen Hochrechnung für den Mikrozensus ab 2005; in: Wirtschaft und Statistik, 8, 739-747.
Kalton,G. (2001): Practical methods for sampling rare and mobile population; in: Proceedings of the Annual Meeting of the American Statistical Association, 454-459
Kohler, U. (2007): Surveys from inside: An assessment of unit nonresponse bias with internal criteria Survey Research Methods, 1, 55-67
Lederer, H.W. (2004): Indikatoren der Migration. Zur Messung des Umfangs und der Arten der Migration in Deutschland unter besonderer Berücksichtigung des Ehegatten- und Familiennachzugs sowie der illegalen Migration, Bamberg
Lavallee, P. (2007): Indirect Sampling, New York:Springer
Löffler, U. (1999): Die Historie der ADM-Stichproben; in: ADM (ed.): Stichproben-Verfahren in der Umfrageforschung, Opladen: Leske + Budrich, 61-68
Marhenke, W. (1997): Telefonanschlußdaten als Auswahlgrundlage; in: Gabler, S./Hoffmeyer-Zlotnik, J. (ed.) Stichproben in der Umfragepraxis, Opladen: Westdeutscher Verlag, 207-220
Opfermann, H./Grobecker, C./Krack-Roberg, E. (2006): Auswirkung der Bereinigung des Ausländerzentralregisters auf die amtliche Ausländerstatistik: in: Wirtschaft und Statistik, 6, 480-494
Promberger, M.(ed.): Neue Daten für die Sozialstaatsforschung. Zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung", IAB-Forschungsbericht, 12-2007, Nürnberg (http://doku.iab.de/forschungsbericht/2007/fb1207.pdf)
Schnell,R. (1991a): Wer ist das Volk? Undercoverage, Schwererreichbare und Nichtbefragbare bei "allgemeinen Bevölkerungsumfragen". In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 43, 1, 106-137
Schnell,R. (1997): Nonresponse in Bevölkerungsumfragen, Opladen
Schnell,R./Kreuter,F. (2000): Das DEFECT-Projekt: Sampling-Errors und Nonsampling-Errors in komplexen Bevölkerungsstichproben; in: ZUMA-Nachrichten, 47, 89-101
Schnell,R./Bachteler, T./Bender, S. (2004): A toolbox for record linkage, in: Austrian Journal of Statistics, No. 1&2., 33, 125-133
Schnell,R./Kreuter, F. (2005): Separating Interviewer and Sampling-Point Effects; in: Journal of Official Statistics, 21, 389-410
Schnell,R./Bachteler,T./Reiher,J. (2005): MTB: Ein Record-Linkage-Programm für die

empirische Sozialforschung; in: ZA-Informationen, Heft 56, 93-103
Schnell,R. (2006a): Alternative Verfahren zur Stichprobengewinnung für ein Haushaltspanelsurvey mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug, Zentrum für Quantitative Methoden und Surveyforschung, Universität Konstanz, unpublished report for the IAB; published in 2007 in: Promberger, M.(ed.): Neue Daten für die Sozialstaatsforschung. Zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung", IAB-Forschungsbericht, 12/2007, Nürnberg, 33-59 (http://doku.iab.de/forschungsbericht/2007/fb1207.pdf)
Schnell,R. (2006b): Ein neues Stichprobenverfahren für einen Haushaltssurvey mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug, Zentrum für Quantitative Methoden und Surveyforschung, Universität Konstanz, unpublished report for the IAB
Schnell,R./Ziniel,S./Coutts,E. (2007): Inaccuracy of Birthday Respondent Selection Methods in Mail and Telephone Surveys; paper presented at the European Survey Research Association Conference, Prague.
Sinn, A./Kreienbrink, A./Loeffelholz, H.D.v. (2005): Illegal aufhältige Drittstaatsangehörige in Deutschland: staatliche Ansätze, Profil und soziale Situation Bundesamt für Migration und Flüchtlinge, Nürnberg: BAMF
Sodeur, W. (1997): Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen; in: ZA-Information, 1997, 41, 58-82
Statistische Ämter des Bundes und der Länder (2004): Ergebnisse des Zensustests; in: Wirtschaft und Statistik, 2004, 8, 813-833
Statistisches Bundesamt (2006): Statistisches Jahrbuch 2006, Wiesbaden