Developments in Data Analysis
A. Ferligoj and A. Kramberger (Editors)
Metodološki zvezki, 12, Ljubljana: FDV, 1996

Different Methods of Survey Sampling in Germany

Jürgen H.P. Hoffmeyer-Zlotnik[1] and Dagmar Krebs[2]

Abstract

In survey research there are mainly two different methods of sampling: first, quota sampling and second, random sampling. Application of quota sampling implies at the first step the construction of a matrix containing all possible combinations of theoretically relevant respondent characteristics being of interest for the individual study. As a second step the target-number of respondents for each combination of characteristics is defined.

In contrast to quota sampling, random sampling is based on probability considerations. A sample is called a simple random sample if each individual in the population is equally likely to be chosen every time we draw an observation where the probability of being chosen has to be > 0. Since we are mainly dealing with face-to-face interviews the paper focuses around sampling for this type of survey research.

In Germany, random sampling for face-to-face interviews is mainly accomplished by two methods: first, drawing a random-route sample on the basis of household addresses listed during interviewer walk in small sample points (random-route sampling); second, drawing a random sample by referring to citizens' addresses filed in the local administration (address-random sampling).

After the theoretical discussion of these different sampling methods two empirical studies are presented and discussed. While the first study compares the demographical sample structure of two representative national samples from the same population drawn (a) by quota sampling and (b) by random-route sampling, the second study compares the demographical sample structure of two representative national samples drawn by (a) random-route sampling and (b) address-random sampling.

[1] ZUMA, Mannheim,Germany
[2] University of Giessen, Giessen, Germany

1. Some theoretical remarks

Sampling for national surveys in Germany using face-to-face interviews is mainly achieved by three different sampling methods:

- quota sampling
- random-route sampling
- address-random sampling.

1.1 Quota sampling

Quota sampling (Koolwijk, 1974; Kromrey, 1980:138; Friedrichs, 1973:133) implies the construction of a matrix containing respondents' characteristics being relevant for the current research question. Characteristics mostly used in quota designs are, for example, gender, age and education. For each cell of the matrix containing a specific combination of respondent characteristics a quota - the aspired number of target persons - is a priori specified. Obviously, there are many possible cells even with a small number of characteristics; for example, for age and education each with three categories and gender (two categories) there are already 18 cells in the matrix of characteristics.. The number of target persons with the cell-specific combination of characteristics has to be proportional to the population distribution of that cells' combination of characteristics (see: Raj, 1968; Konijn, 1973:189).

Thus, from the point of view of the theoretical assumptions, the design of a quota sample is very similar to a simple random sample. But while in random-routes sampling and especially in address-random sampling the rules for contacting a specific person are specified, this is not the case in quota sampling. Here the interviewer has a big amount of freedom to choose a specific person for an interview as long as she/he sticks to the combination of characteristics a priori defined by the „quota-matrix" (Noelle-Neumann, 1963). Thus, the researcher usually knows little about the strategy applied by interviewers to complete the interviews with the a priori specified number of respondents possessing the prescribed quota characteristics. The interviewer is free in selecting a specific person as respondent. The researcher mainly expects that the number of interviews with respondents possessing the target characteristics corresponds to the a priori specified number and that the interviews are completed within an a priori defined sample point. In the ideal case, each interviewer should be restricted to only one sample point.

The PSU's (Primary Sampling Units) serving as sample points are constructed by sequentially proceeding from big to small units applying regionally stratifying variables:

- (federal) state
- county

- cities classified into size-types, generally:
- big town, middle town, small town, village
- with big towns differentiated into:
- zones or quarters or neighborhoods

Therefore, in the case of a villages or a small town a PSU (a sample point) contains the whole village/small town while in middle and big towns the sample point refers to a specified part of the town. Thus, with respect to regional stratification and regional distribution the comparability of a quota sample to a random sample can be controlled by selection of sample points and the a priori specification of the number of target persons (possessing the prescribed combination of characteristics) to be interviewed.

Table 1.1: Schematic Display of Quota Sampling

1. Step:
   Stratification of the population into state, county and towns classified with respect to the number of inhabitants into groups of community size; stratification of the population in big towns into part of the town
   Systematic selection of sample points proportional to the predefined strata

2. Step:
   Construction of a matrix containing the combination of respondent characteristics associated with the number of target persons (the quota design)
   Defining the number of interviews to be completed by each interviewer as well as in each sample point

Result:
   No random sampling

With respect to the freedom of interviewers to select respondents quota sampling has often be blamed for reflecting the network of interviewers but not being a random sample.

1.2 Random sampling

In contrast to quota sampling, random sampling implies a systematic selection based on probability criteria. The fact that by systematic random sampling (see: Levy and Lemeshow, 1991:43; Raj, 1968:61) each „element" in the population is (theoretically) equally likely (with a probability > 0) to be chosen into the sample is considered as an important difference to quota sampling.

The principle of random sampling (Kish, 1965) is shortly described with respect to the „bowl-model": Simple random sampling is based on this model assuming that all „elements" of the population, for example addresses, are contained in this bowl. After carefully shuffling, as many „elements" as needed for the sample are at random chosen. In each single choice each „element" of the population is equally likely to be chosen.

In stratified random sampling (Hansen, Hurwitz, and Madow 1953; Kish, 1965) the first step is to divide the population with respect to relevant characteristics into several different strata. In a second step, within each stratum a simple random sample is drawn in such a way that each „element" is equally likely to be chosen into the sample. If simple stratifying does not correspond to the research problem, a sampling method encompassing different stages has to be applied. In this context one can think of a method combining different sampling procedures, for example a procedure where the strata are defined as „elements" of the population. At each stage in such multi-stage sampling one has to be aware of dealing with a sub-quantity of the former stage (Kish, 1965; Sudman, 1976). Drawing a national sample by random sampling implies multi-stage stratified sampling leading in the end to a selection of a manageable number of PSU's.

In random-route sampling the PSU's are small units of homogeneous quarters like "voting districts" with at least 400 but in the average 2000 potential target persons (usually over 18 years old with German citizenship living in private households).

In address sampling the PSU's are typical settling areas like villages or towns or quarters (sections of the city).

1.2.1 The random-route sampling design

Random-route sampling contains several sampling procedures resulting in different sampling designs. The sampling design mostly used with random route sampling in Germany is the ADM-design deriving its name from the "Working Group of German Marketing Research Institutes", the so called ADM. This sampling design is a multi-stage stratified random procedure (Arbeitsgemeinschaft ADM-Stichproben & Bureau Wendt, 1994; Schaefer, 1979).

In the ADM-design sample points are voting districts. 210 of these voting districts are taken together to constitute a so called net. All sample nets are free of overlapping and are representative of the national population. Thus, for studies needing a large number of cases several of these nets can be combined without damaging the national representativity of the sample as a whole (see: Kirschner, 1984).

The selection of sample points is based on systematic random sampling from a structured sequence of voting districts. This structured sequence has been constructed according to regionally stratifying variables (Arbeitsgemeinschaft ADM-Stichproben and Bureau Wendt, 1994:194):

- 1. stage: federal states
- 2. stage: within each federal states districts
- 3. stage: within each district counties
- 4. stage: within each county town-types
- 5. stage: within each town-type (from middle town upward according to size regarding the number of inhabitants) specifying communities, urban zones and voting districts and ordering these units sequentially according to a weight variable displaying the number of households in that unit.

The selection interval for drawing a sample of PSU's is defined as the ratio of the total number of households in the population and the total number of voting districts in the country. The starting point is defined as a random number between 1 and the length of the selection interval (Arbeitsgemeinschaft ADM-Stichproben & Bureau Wendt, 1994:194;. Kirschner, 1984:121).

Arriving at this level of sample points (the voting districts selected in the former step), not all persons („elements") of the population living in the selected voting districts can become respondents in a current survey sample. Therefore a sample of target households is drawn within the sample points by random-route sampling. The procedure is as follows: each interviewer gets a starting address within a defined sample point. According to specific walking prescriptions the „random walk" of the interviewers starts from this address. The walking rule prescribes to list the address of every third private household on a specially prepared formular up to a maximum of 69 households. Units not belonging to the population of private households, for example shops, offices, institutions like hospitals. This list of addresses in the sample point defines the total of units (households) eligeable for the current survey sample. From this address listing every eighth address is selected. The resulting number of addresses constitutes the gross sample of households to be contacted. By fixing the number of addresses to be listed on the prepared formular a response rate of 65-70% is assumed.

Starting with this gross sample of household addresses the interviewer has to select exactly one person (of the a priori defined population) per household as respondent. This selection is structured by a systematic procedure: first, all persons (belonging to the population of interest) in each household of the gross sample are listed according to age in decreasing order; second, the target person serving in the end as respondent is selected according to a list of random numbers individually attached to each household. For example, from a household containing i persons belonging to the population of interest a respondent is selected with a probability of 1/i. Thus, the order of permutation is constructed in a way that the likelihood of being chosen is 0.5 for a person in a household with two persons, 0.33 in a household with three persons and 0.25 in a household with four persons.

To summarize (Arbeitsgemeinschaft ADM-Stichproben & Bureau Wendt, 1994: 194;. Kirschner, 1984): The describes random-route sampling leads in a first step to a sample of PSU's. Are the PSU's considered as given, random-route sampling results

in a second step to a sample of households where each household is equally likely to be chosen. In a third step in each household, regardless of the number of persons in the household, one and only one respondent is selected. This describes the disproportional likelihood of a person to be selected as respondent in a current survey: persons from big households have a smaller chance to be selected than persons from small households (1/i). The disproportionality in the likelihood of being selected into the current sample has to be compensated for by a weighting variable (the number of households is multiplied by the number of persons per household and the product is standardized), the so called „design weight".

Table 1.2: Schematic Display of the Random-Route Design

1. Step:
   Stratification of the population according to state, district, county and community type (classified according to number of inhabitants)
   Systematic selection of sample points proportional to strata with random starting point

2. Step:
   Within each sample point selection of an equal number of private households containing at least one person belonging to the target population
   systematic random selection
3. Step:
   Selection of one target person (respondent) per household
   systematic random selection

Result:
   Selection probability for each respondent is inversely proportional to household size

therefore:
4. Step:
   Weighting

1.2.2 Address-random sampling

Address sampling is the second way of getting a stratified random sample of respondents. Again one starts with constructing regional strata leading in the end to sample points as the smallest regional units: villages or small towns or quarters of big towns. However, drawing a sample of choosing sample points in address random sampling implies a proportional selection of loci according to a typology of state, district and county. This typology includes in the first step villages, small towns and big towns. In a second step sample points within big towns are chosen at random

from a typology of quarters of the big towns. Careful attention should, however, be directed to limiting the number of sample points because each community has its own rules in address handling resulting in an organisational mess with a large number of sample points.

Address sampling in this form is possible in Germany where each inhabitant is registered by the administration of her/his place of living. Thus, in each sample point all inhabitants possessing the characteristics of the target population (cohort membership, gender or ethnic group) can be identified. According to the structure of the files which the sampling procedure has to take into account (i.e. whether the names are sorted alphabetically or by addresses) a sample of persons can be drawn. The number of persons to be included into the sample depends on the number of persons possessing the characteristics of the target population.

Table 1.3: Schematic Display of Address-Random Sampling

1. Step:
   Stratification of the population according to state, district, county and community type (number of inhabitants classified according to size)
   Systematic selection of sample points proportional to town-size or type of settlementing big cities selection of sample points proportional to type of section/zone of city

2. Step:
   In each sample point selection of a number of persons (possessing the characteristics of the target population) proportional to the total number of potential target persons in that sample point.
   Systematic random selection

Result:
   Equal likelihood of being chosen for each respondent from the target population

1.3 Summary

Comparing the three sampling procedures leads to the conclusion that:

- quota sampling leaves the factual selection of respondents from the theoretically prescribed „quota" completely to the interviewer.
- random-route sampling gives a strict prescription of getting over several steps to the target person. If these prescriptions are taken seriously, different interviewers must end up with identical respondents. Nevertheless, this

sampling procedure still leaves it to the interviewers to „find" the target person. Although control of interviewers is stronger in this sampling design, there still are some degrees of freedom left.

- address-random sampling exactly determines target persons by name and address. There are no degrees of freedom left for the interviewers and optimal control is guaranteed.

With respect to the costs of time and money the three sampling procedures can be characterized as follows:

- quota sampling is fast to accomplish and inexpensive. Since the selection of target persons is left to the interviewers, sampling does not require time and money.
- random-route sampling requires the random walk, listing of household addresses as well as listing of persons to define the target person. Thus, the procedure takes two to three times more time and consequently more money to establish the desired sample.
- address-random sampling requires even more time (in Germany five months are not unusual) to establish the sample. Compared to random-route sampling the amount of money for this sampling procedure can be about 30-50% higher because of the money the researcher is charged by the administration for the addresses.

Finally, with respect to the likelihood of being chosen the two random sampling procedures differ. While random-route sampling (in principle based on a sample of households) endows persons from smaller households with a higher likelihood of being chosen into the sample (according to the rule of random walk: one person per household), address-random sampling (based on a sample of persons) persons from bigger households have a higher likelihood of being chosen into the sample (according to selection independent of household membership). Regardless of the costs, some researchers prefer the latter method because of the household-independent sampling.

2. Quota sampling versus random-route sampling

2.1 Comparison of two studies

Compared are two national surveys referring to nearly identical target populations: first, the youth study of the Shell Foundation conducted during early summer of 1991 and the youth survey of the German Youth Institute conducted during fall of 1992.

The sample of the Shell youth study was based on quota sampling containing N=4005 young people between 13-29 years of age in both parts of Germany; 66.6% in West and 33.4% in East Germany. In the Western part of Germany interviews were conducted in 410 sample points, in the Eastern part in 320 sample points being villages, small towns or quarters of big towns. Within big towns a reasonable regional dispersion of interviews was accomplished by fixing the average number of interviews per sample point to 4-6. For example, the addresses of respondents living in the 10 biggest East German towns are distributed over about 100 CIP-Codes. The CIP-Code system in Eastern Germany gives different numbers to different quarters of the town, that is, that in the 10 biggest East German cities a reasonable dispersion across the city area could be accomplished.

The sample of the youth survey of the German Youth Institute (DJI) was based on random-route sampling corresponding to the ADM-sample design containing N=7009 young people between 16-29 years of age; 63.8% of the interviews were done in West, 36.2% in East Germany. All together 1.470 sample points were used, 945 in West and 525 in East Germany. A sample point in this sample design represents a voting district. The response rate was 65.5% in West and 66.2% in East Germany.

The two sampling procedures have in common the first step - selection of the regional unit "village" or "town" (small, middle or big) - based on census data describing the regional distribution of the population. However, in contrast to quota sampling random-route sampling defines sample points as homogeneous neighbourhoods. according to the description of random-route sampling (above) interviewers have theoretically no degrees of freedom in selecting respondents.

This, however, is not true for quota sampling. As soon as the regional units are fixed, interviewers are completely free to select respondents as long as they follow the prescribed combination of the given characteristics and select the required number of persons. In the Shell youth study these characteristics were: age with 5 categories, gender with two and education with 3 categories. Thus, the matrix of combination of characteristics consisted of 30 cells (see Table 2.1).

Since in quota sampling interviewers are completely free not only in recruiting respondents but also in the context where they recruit them - for example in their own social networks, in their neighbourhood, before a youth club, a school or a university - it is crucial that the group of interviewers displays a cross-section of all population groups; this requirement is primarily necessary to prevent bias in the sample.

The other side of the coin is the number of interviews to be completed within a regional unit, for example a town. Usually 4-6 interviews are recommended per sample point. The number of sample points, however, depends on the density of the target population in that regional unit. While in random-route sampling a big regional unit contains a large number of sample points with a comparatively small number of interviews to be completed, in quota sampling the same regional unit contains a

smaller number of sample points with a comparatively large number of interviews to be completed.

Table 2.1: Matrix of Characteristics of the Quota-Design

		Gender:	
Age	Education*)	Male	Female
13-15	low		
	middle		1
	high	1	
16-18	low		1
	middle		
	high		
19-21	low		
	middle		
	high		
22-24	low	2	
	middle		
	high		1
25-29	low		
	middle		1
	high		

*) for people still in education: the aspired educational level

2.2 The specific case of two youth studies

The Shell youth study as well as the DJI youth survey contained comparable birth cohorts 1962-1976. The distributions of these cohorts are compared to that of the micro-census data. The census data are shown in table 2.2.

Two national surveys aspiring to be representative of the target population are expected to display comparable sample structures. This is not the case here because in the Shell youth study whole communities served as sample points. In quota sampling, within these communities interviewers are free to choose specific residential quarters according to criteria, (here for example, the criterium that many young people of the target population live in a specific quarter). These specific residential quarters are not defined by the sampling design but are arbitrarily chosen by interviewers. In contrast, in random-route sampling sample points are defined as residential quarters (homogeneous neighborhoods) according to criteria not referring

to the target population (here for example young people) but to the total residential population.

Table 2.2: Expected Distribution of Birth Cohorts in both Youth Studies According to Micro-Census Data

Birth Cohorts	West	East
76-73	21.1	22.5
72-70	18.5	18.8
69-67	24.4	21.4
66-64	26.3	27.3
63-62	9.7	10.0

Table 2.3: Observed Distribution of Birth Cohorts (Unweighted) Obtained from the SHELL Youth Study and the DJI Youth Survey

	SHELL		DJI	
Birth Cohorts	West	East	West	East
76-73	27.6	28.5	22.7	29.2
72-70	24.8	27.2	17.9	19.8
69-67	24.8	25.0	23.1	19.3
66-64	15.2	13.3	26.3	22.7
63-62	7.6	5.9	10.0	9.0

In general, the random-route sample fits the (from the micro-census data) expected distribution better but there is still a discrepancy between what „should be" and what one „gets". Interviewers have a maximum of four trials to contact a target person and to complete an interview (see Table 2.3) and thus limits the likelihood to reach all by the sampling procedure chosen persons. Nevertheless, with a sample based on random-route sampling it is possible to contact even the highly mobile (and therefore difficult to reach) persons being in the middle of their twenties. The advantage of listing household members is obvious especially with respect to this mobile group of persons: target persons become visible and can be "prosecuted" or "traced" by different trials to contact them. In quota sampling, however, there is no specific person to trace. Instead a specific combination of characteristics has to be found. Thus, primarily persons being easy to reach or being cooperative are interviewed. Since a target person is usually not contacted in her/his household these interviewed persons can - according to the interviewer's strategy of contacting - belong to a specific subculture (her for example visitors of a disco, a school, a cinema etc.). Therefore the discrepancy between a sample based on quota-sampling

and the (from the micro-census data) expected distribution must be bigger than that of a sample based on random-route sampling.

Another factor effecting a structural difference in national survey samples between the random-route and the quota sample results from the relation between location of sample point and residential position of interviewers. An institute working with the ADM-sample design buys (from the ADM group) sample nets with regionally fixed sample points. Around these sample points interviewers are recruited. In general, in big cities nothing changes in this structure if new sample nets are bought because big cities will always be represented in at least some of the sample nets. Problems occur, however, in rural areas because first, few interviewers live in rural areas, and second, interviewing in rural areas implies covering long distances. If an institute does not use a priori defined sample points interviewers are recruited at locations where their activity is most efficient. This is the case in urban areas. (In the beginning of the nineties this was not the case in the Eastern part of Germany because interviewer organisations of big research institutes did not exist there before unification.)

Table 2.4: Distribution of Community Type (Classified by Number of Inhabitants) in Expected (DJI Weighted by Census Data) and Observed Distributions of the Shell Youth Study and the DJI Youth Survey

	Expected: DJI Weighted		Observed: DJI		Observed: SHELL	
Community Type	West	East	West	East	West	East
below 2000 inh.	5.5	23.2	5.0	23.6	7.8	15.1
2- 5000 inh.	8.3	9.5	7.9	10.0	10.0	13.8
5- 20000 inh.	26.1	15.2	22.0	15.8	20.1	21.1
20- 50000 inh.	16.3	15.0	14.6	15.3	12.4	12.7
50-100000 inh.	9.3	8.0	8.9	7.6	10.0	9.9
100-500000 inh.	17.9	16.8	19.0	15.8	20.5	12.7
above 500000 inh.	16.5	12.3	22.6	11.8	19.2	14.8

As the data in Table 2.4 show respondents from small villages in East Germany are significantly underrepresented in the quota sample compared to their representation in the random-route sample. The bias related to this small response rate in villages is still increased by the comparatively high (compared to what „should" be) response rates realised in small towns. The hypothesis that an over-sampling of respondents in big (West German) towns occurs especially with quota sampling is not confirmed by the data in Table 2.4. However, an unexpected over-sampling is observed in the random-route sample in big (West German) towns. This is remarkable because in general, response rates in national random-route samples are low in big towns. This unexpected result might be a consequence of the special effort interviewers working in big towns usually have to perform.

Table 2.4 displays a classification of settlement according to number of inhabitants. In Table 2.5 this classification is broken down by those federal states having different settlement structures. Additionally, East and West Germany have a different infrastructure with respect to roads and highways. Covering long distances on small roads being in poor condition implies the cost of time. Investments of this kind made by interviewers are usually not adequately payed by the institutes. Although this fact applies more to quota sampling than to random-route sampling (in the latter distances between interviewers' residence and sample points are a priori known) it is effective in both sampling designs. The differential characteristics of the infrastructure thus influence the behavior of interviewers leading to differential response rates in areas difficult to reach.

For example, in the states Baden-Württemberg und Bayern rural areas are well covered by highways and can easily be reached by interviewers. Thus, with respect to sampling design, one would not expect differential response rates in rural areas or villages. Inspection of the contact protocols (of interviews completed under the quota sampling design), however, reveal first, an unexpected and economically unreasonably high percentage of interviews completed in small villages (mostly only one or two interviews). Second, these villages are usually located in the periphery of cities. Taking into account this information one can suspect - especially for the target population of juveniles - that many respondents coming to the cities for example, on weekends to visit a disco, a youth club etc. are contacted and interviewed there.

In Brandenburg, however, another picture is observed. A high percentage of this state's population lives in small villages with less than 500 inhabitants. These villages are difficult to reach by highways or public transportation. With the exception of the surroundings of Berlin, there are practically no urban areas. Thus, young people from small villages are not easily reached even if they visit discos and youth clubs because these institutions are also difficult to reach. Thus, as was shown above, communities with 2.000-5.000 inhabitants are over represented in the Eastern part of the country in the quota sample. In the special case of the Brandenburg sample part, most small towns belong to the direct surroundings of Greater Berlin.

With respect to respondents' characteristics there are some minor differences between the two samples especially in rural areas. Respondents being still in school or educational institutions as well as respondents of younger age groups are slightly over-sampled while housewives are slightly under-sampled in the quota sample compared to the random-route sample. Referring to the above discussed infrastructural conditions in some federal states this under representation of housewives might be ascribed to the socio-demographic structure of disco and youth club visitors: housewives usually do not belong to the regular population of disco visitors. Additionally, this result also reveals again the quota interviewers' strategy of recruiting target persons: usually these interviewers do not contact private households.

Table 2.5: Distribution of Community Types (in Thousand) in Quota (QT) and Random-Route Design (RR) in Selected States

Community Type	- 2000		2000-5000		5000-20000		100000-500000		500000 +	
sample	QT	RR	QT	RR	QT	RR	QT	RR	QT	RR
State										
Schleswig-Holstein	18	22	8	13	24	24	23	16	-	-
Niedersachsen	9	8	14	5	38	30	14	17	2	13
Nordrhein-Westfalen	2	-	2	-	14	10	32	34	22	24
Baden-Würtemberg	6	1	13	15	24	32	26	14	3	8
Bayern	14	7	13	20	26	30	14	12	19	16
Sachsen	15	21	14	10	24	16	13	17	18	12
Brandenburg	11	31	23	9	28	23	6	8	2	-

2.3 Summary

Although there are some differences between the micro-census and the quota sample distribution in the rural area, the latter reproduces the target population quite well. This can be ascribed to the big number of sample points providing a good dispersion. It has been suspected that in quota sampling interviewers contact target persons in group specific infrastructural settings like discos etc. Whether this strategy results in an under representation of the heterogeneity of attitudes and/or opinions cannot be discussed here. Those attitude questions being comparable for both surveys are related to politics and society in the unified Germany. For these attitudes the one year's distance between the two surveys completely destroys any comparability.

3. A Comparison of random-route sampling versus address-random sampling

3.1 Study description

In 1988 the DJI conducted a survey on "Partnership and Family Today". The target population consisted of persons between 18-55 years of age living in private households. From this population a sample of N=9942 was drawn by two different sampling procedures: first, N=3011 persons were chosen by address-random sampling from file (address-random); second, N=6931 persons were chosen by random route sampling (random-route). Both procedures should result in representative samples which reflect the structure of the target population. Thus, two questions are primarily discussed:

1. Are the distributions of socio-demographic characteristics realised in the address-random sample comparable to those realised in the random-route sample?
2. To what degree are differences between the two samples due to the behavior of interviewers?

First, however, the question of response rate in both samples is described together with the reasons for non response.

3.2 Response rate

The structure of non response in both samples is different. Starting with the "cleaned total" (the sample size resulting after substracting the so called "neutral non response" occurring before contacting the target households (random-route sampling) or target persons (address-random sampling)) the reasons of non response are displayed in Table 3.1.

These numbers reflect mainly two striking aspects: first, the higher refusal rate in the address-random sample and second, the higher amount of households not being reachable for the interviewers in the random-route sample.

These results indicate that the random-route sample is characterized by a higher (than the address-random sample) percentage of cooperative persons (being generally willing to give an interview) while at the same time it is characterized by a higher percentage of households not reachable for the listing necessary for the selection of the target person. While the former results from the sampling procedure of contacting in the first step the household (if the household is cooperative, the selected target person usually complies), the latter results from the higher percentage of one-person households in the random-route sample (see 1.3): one-person

households (singles) are the typical highly mobile persons who are difficult to reach. Thus, in random-route sampling, agreement of a household to the listing of its members by the interviewer indicates a general cooperative attitude of the household which usually is transferred to the target person. In contrast, in being directed to a specific person instead of being directed to a household, address-random sampling does not have this kind of support of the household and thus contains a higher rate of „general refusals".

Table 3.1 Reasons For Non Response Based On Cleaned Total In Address-Random Sample (AR) And Random-Route Sample (RR).

	AR		RR	
	Abs.	%	Abs.	%
Cleaned Total	5489	100.0	13733	100.0
Realised Interviews	3011	54.9	6931	50.5
Reasons for Non Response:				
Nobody Reached in Household	229	4.2	1856	13.5
Target Person: Not Reached	225	4.1	456	3.3
Target Person: Out of Town	73	1.3	291	2.1
Target Person: Ill	88	1.6	134	1.0
Target Person: No Time	278	5.1	1325	9.6
Target Person: General Refusal	1560	28.4	1668	19.4
Other Reasons	19	0.4	28	0.2
Incomplete Interviews	6	0.1	44	0.3

To summarize: the realised random-route sample differs from the realised address-random sample not only in the structure of the reasons for non response but also in one important aspect: respondents in the random-route sample are more easily reached than respondents in the address-random sample because the singles who are difficult to reach are no longer contained in the realised random-route sample.

3.3 Number of contact trials and realized interviews

Researchers (Allerbeck and Hoag, 1981; Hoag, 1993) sometimes tend to ascribe the structural characteristic of the random-route sample - over representing the "easily reachables" - to specific interviewer strategies using the degrees of freedom built into random-route sampling. The logic of the two sampling procedures in principle implies that persons drawn by the random-route procedure are more difficult to reach than persons drawn by the address-random procedure because the latter are known by name and address while the former have to be "digged out" during the selection

process. In reality, however, this is not the case as the data in Table 3.2 clearly demonstrate.

Nearly 50% of the interviews of the random-route sample are completed during the first contact with the household from which the target person has to be selected. In contrast, only one fourth of the interviews of the address-random sample are completed during the first contact. This seems to indicate - and thereby support the suspicion - that interviewers manipulate the selection rules of the random-route sampling: if the target person selected by the random-route sampling procedure is not available the interviewer might select another person as respondent who also belongs into the target population. Or the interviewer changes to another household if nobody answers in the target household prescribed by the random-route selection rules.

Table 3.2 Number Of Contacts And Completed Interviews In Address-Random Sample (AR) And Random-Route Sample (RR)

	Percentage of Interview Completed at				
	First Contact	Second Contact	Third Contact	Fourth Contact	Fifth and more Contact
AR	25.0	34.5	18.9	10.8	10.8
RR	44.2	33.7	14.6	5.5	2.0

This "optimizing" strategy can results in an overrepresentation in the random-route sample of those persons who are easy to reach and thus effect structural differences between address-random and random-route samples.

3.4 Demographic structure of random-route sample and address-random sample

The following distributions result after weighting the data of the random-route sample by the design weight (see 1.3). After weighting the demographic structure of both samples should be approximately equal. Table 3.3 gives the distribution of selected demographic characteristics equally available for both samples. Besides characteristics of respondents indicators of housing and region are given.

Persons who constitute the group of „difficult to reach" persons are not contained in the sample. Following the former discussion, these are singles living in one-person households. Additionally, occupied persons also belong to this category. In the random-route sample persons who are difficult to reach belong to the category of non responses being excluded in the process of sampling. Thus, one important

difference between address-random and random-route sample is imbedded in the sampling procedure: by excluding the persons who are difficult to reach, the random-route sample contains a higher percentage of „easily reachable" persons than the address-random sample.

Table 3.3 Demographic Characteristics in Address-Random Sample (AR) and Random-Route Sample (RR)

	AR	RR
Gender		
Male	50.1	43.3
Female	49.9	56.7
Occupational Status		
Occupied	68.2	60.5
Double Income No Cid	57.9	55.6
Age		
18-25	21.0	20.4
26-35	28.4	27.8
36-45	24.0	25.3
46-55	26.5	26.5
Education		
Compulsory Schooling	49.7	47.4
Post Comp. Schooling	34.7	34.4
Graduate or Higher	15.6	18.1
Marital Status		
Married	64.1	68.4
Divorced	5.1	4.6
Widowed	1.5	1.4
Not married	29.3	25.5
Married Women Without Occ.	15.9	22.8
Housing		
House	43.6	43.0
Flat/Appartment	42.0	46.3
Room/Parents	14.4	10.6
Settlement Type		
Below 20.000 Inhabitants	46.2	39.9
20.000 - below 100.000 Inhabitants	25.0	27.0
100.000 and more Inhabitants	28.8	33.8

The first information from Table 3.3 is that the ratio of females to males differs considerably between the random-route and the address-random sample: the gender difference in the random-route sample is 13% while in the address-random sample both gender groups are nearly equally represented. Second, the random-route sample contains more (4.3%) married persons than the address-random sample and it also over represents - with a difference of 6.9% - married women without occupation. With respect to occupation the random-route sample includes 7.7% fewer occupied persons than the address-random sample but represents DINCs (married occupied double income persons) in nearly the same way (the difference is 2.3%) as the address-random sample. With respect to age the two sampling procedures do not result in different distributions. The same holds for education, differences in percentages being maximal 2.5%. Differences are, however, observable for type of housing and settlement type classified according to number of inhabitants: persons in the random-route sample tend to live more often in appartments and in big towns while in the address-random sample persons more often live in rooms and in small towns.

If these demographic characteristics are classified with respect to easy reachability one can conclude that easy reachability affects the sample distributions. Compared to the address-random sample the random-route sample over represents females, persons without occupation, married persons, housewives, persons living in flats within big appartment houses, persons living in big towns. With respect to „easy reachability": in big towns it is much easier to reach addresses because there are more households concentrated on smaller space.

3.5 Summary

While there are differences between the address-random and the random-route sample with respect to demographic characteristics, these differences are mostly small. Nevertheless, these differences result mainly from the principally different sampling procedures in address-random and random-route sampling design. Resulting from the sampling procedure distributions of address-random samples come closer to those of reference data like micro-census distributions than random-route samples. To decide, however, which of the two sampling procedures is the best, one has to consider not only the differences in sample distributions but also take into account the costs involved with either sampling procedure. Where differences occur they altogether do not reach the 10% level. The decision for or against address-random sampling thus is reduced to the availability of monetary resources. As describes in this paper, costs for address-random sampling exceed those for address-random sampling by about 50%. Thus, the recommendation is to improve the quality of random-route sampling by separating household listing from interviewing thereby reducing the degrees of freedom for the interviewers in selecting the target person. Introducing this procedure into random-route sampling would result in an increase

the costs for an random-route sample by about 20% compared to the costs of presented random-route sampling procedures.

4. Conclusion

The three presented sampling procedures differ mainly with respect to sampling theory as well as expenses of time and money in sampling itself. Differences primarily exist between quota sampling and random sampling.

First, the likelihood of inclusion can not be determined in quota sampling but is theoretically calculable in random sampling. Second, random-route sampling proceeds in consecutive steps selecting first the household and from the household the target person thereby implying that a person's likelihood of being chosen as target person is inversely proportional to household size.

However, the comparison of real sample distributions based on quota as well as random sampling, reveals that differences of sampling procedures tend to level either by definition of sample points, prescriptions for random-walk or for selection of addresses from file, or high non response rates.

Crucial for the quality of a sample is a regional stratification covering a maximum of sample points optimally distributed over all existing types of settlement in all regions. Less crucial but still important for the quality of a sample is the minimization of degrees of freedom for interviewer decision in selecting target persons. As far as these criteria are carefully observed, even quota sampling can be better than its reputation and address-random sampling can loose importance because is imbedded expenses of time and money.

References

[1] Allerbeck K. and W. Hoag (1981): Interviewer- und Situationseffekte in Umfragen. Eine log-lineare Analyse. *Zeitschrift für Soziologie*, **10**: 413-426.

[2] Arbeitsgemeinschaft ADM-Stichproben & Bureau Wendt (1994): Das ADM-Stichproben-System (Stand 1993); in: S. Gabler, J.H.P. Hoffmeyer-Zlotnik and D. Krebs (Eds.), 188-202.

[3] Arbeitskreis Deutscher Marktforschungsinstitute ADM (Eds.), bearbeitet von F. Schaefer (1979): Muster-Stichproben-Pläne für Bevölkerungs-Stichproben in der Bundesrepublik Deutschland und West-Berlin. München: Verlag Moderne Industrie.

[4] Cochran W.G. (1972): *Stichproben-Verfahren*. Berlin: De Gruyter.

[5] Friedrichs J. (1973): *Methoden empirischer Sozialforschung*. Reinbek: Rowohlt.

[6] Gabler S., Hoffmeyer-Zlotnik J.H.P., and D. Krebs (Eds.) (1994): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag

[7] Gonzales M., Ogus J., Shapiro G., and B. Tepping (1975): Standards for Discussion and Presentation of Errors in Survey and Census Data. *Journal of the American Statistical Association*, **70**: 7-15.

[8] Groves R. M. (1989): *Survey Errors and Survey Costs*. New York: John Wiley.

[9] Hajek J. (1981): *Sampling from a Finite Population*. New York: Marcel Dekker.

[10] Hansen M.H., Hurwitz W.N., and W.G. Madow (1953): *Sample Survey Method and Theory*, Vol. 1/2. New York: John Wiley.

[11] Hartmann P. (1990): Wie repräsentativ sind Bevölkerungsumfragen? Ein Vergleich des ALLBUS und des Mikrozensus. *ZUMA-Nachrichten*, 26: 7-30.

[12] Hess I. (1985): *Sampling for Social Research Surveys, 1947-1980*. Ann Arbor: The University of Michigan Press.

[13] Hoag W. (1993): Zur Qualität der Daten der Baseline-Studie 1991 & ALLBUS 1992. Unpublished paper.

[14] Kaplitza G. (1975): Die Stichprobe; in: Holm, K. (Eds.): *Die Befragung* 1: 136-186. München: UTB 372.

[15] Kirschner H.-P. (1984): ALLBUS 1980: Stichprobenplan und Gewichtung; in: Mayer K.U. and P. Schmidt (Eds.): *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*. ZUMA-Monographien Bd. 5: 114-182. Frankfurt/M.: Campus.

[16] Kish L (1965): *Survey Sampling*. New York: John Wiley.

[17] Konijn H.S. (1973): *Statistical Theory of Sample Survey Design and Analysis*. Amsterdam: North Holland.

[18] Koolwijk J. van (1974): Das Quotenverfahren; in: Koolwijk & Wieken-Mayser (Eds.): *Techniken der empirischen Sozialforschung* Bd. 6: 81-99. München: Oldenbourg.

[19] Kromrey H. (1980): *Empirische Sozialforschung*. Opladen: Leske und Budrich.

[20] Lessler J.T. and W.D. Kalsbeek (1992): *Nonsampling Error in Surveys*. New York: John Wiley.

[21] Levy P.S. and S. Lemeshow (1991): *Sampling of Populations: Methods and Applications*. New York: Wiley.

[22] Mendenhall W., Ott L. and R.L. Schaeffer (1971): *Elementary Survey Sampling*. Belmont CA.: Duxbury Press.

[23] Noelle-Neumann E. (1963): *Umfragen in der Massengesellschaft*. Reinbek: Rowohlt.

[24] Raj D. (1968): *Sampling Theory*. New York: McGraw-Hill.

[25] Sirken M.G. (1970): Household surveys with multiplicity. *Journal of the American Statistical Association*, **65**: 257-266.

[26] Sudman S. (1976): *Applied Sampling*. New York: Academic Press.

[27] Thompson S.K. (1992): *Sampling*. New York: John Wiley.