Sample design and consequences

Schubert, Peter; Greil, Angelika

CHAPTER 2

SAMPLE DESIGN AND CONSEQUENCES

PETER SCHUBERT AND ANGELIKA GREIL

2.1 Introduction

In chapter 1 it was indicated that in this book the results of two different approaches to the collection of survey data will be compared. One study, the standard Eurobarometer 41.0, used face to face interviews in 12 EU member states and was carried out by the research company INRA which normally collects the data for the Eurobarometers. The second study used as data collection method telephone interviews and has been carried out by FORSA in the same 12 EU member states. The purpose of this study was to assess the feasibility of telephone interviewing in the EU member states. By comparing the results of these two studies one can see how large the differences will be between the results of a face to face and a telephone survey in Europe. For methodological purposes the standard Eurobarometer study was augmented by a panel component in form of a telephone study also done by INRA in three countries in order to study the mode effects of the different approaches on a series of questions, the countries being Belgium, Spain and France.

Since the differences in sampling design and fieldwork in the three studies can have caused differences in the results, these differences in design and fieldwork will be described. After that the weights to correct for differences from the populations will be discussed, and finally a comparison of the response rates in the studies will be given in order to look at some fieldwork-related differences.

2.2 Fieldwork

Fieldwork of all three surveys took place in the spring of 1994. The populations to be studied in the face to face and in the telephone study were identical: the populations of the 12 EU member states aged 15 years and older. The target sample of the INRA surveys was 1000 interviews per country, and 500 interviews per country with the FORSA surveys. The INRA face to face - telephone panel study aimed only at a sample of 3 countries because of its special purpose.

2.2.1 The face to face study: The standard Eurobarometer 41.0

The Eurobarometer 41.0 (EB41.0) survey was carried out by specialised polling firms co-ordinated by INRA Europe, Brussels. From April 4th until May 6th, 1994, the respondents were contacted in their private homes and questioned in face to face interviews.

In order to select the respondents a stratified multi-stage random sample routine was used. The sampling frame for EB41 were the smallest enumeration units of the census in each country. Of these units a random sample was drawn stratified on the basis of the regions (Eurostat NUTS II level) and degree of urbanisation. In this way at least 100 primary sampling units were selected in each country. Within each sampling unit a starting address was drawn in a random fashion from an official address book or otherwise. In most countries more than 100 starting addresses were chosen. A randomly chosen increment determined the 10 more address per sampling unit in order to obtain a sample of approximately 1000 households. Within a household, the actual respondent was selected among those aged 15 or older by the criterion of who in the household had the next birthday or the Kish-grid.

The training of the interviewers was the responsibility of the national member of INRA. There are no agreed-upon supervision procedures between the national representatives of INRA except the general rule to visit each household only one more time if at the first time no interview could be done. After two failures the household was registered as nonresponse. Respondents who refused to co-operate have not been contacted once more for conversion purpose. If at a household no interview could be done, this address was substituted by a random walk procedure.

Table 2.1 Completion rate EB41.0

EB41.0	B[2]	DK	W-G	E-G	GR	ESP	F
Gross sample	2318	2739	2358	2208	2007	2354	2190
Net sample	1087	1005	1064	1058	1010	1003	1034
Completion rate (%)	47	37	45	48	50	43	47

	IRL	NIRL	I	Lux	NL	P	GB	EU12
Gross sample	2025	944	2636	1447	2632	1975	2155	29988
Net sample	1068	306	1058	625	1015	1002	1067	13402
Completion rate (%)	53	32	40	43	39	51	50	44.7

Table 2.1. displays the completion rate of the Eurobarometer 41.0 in the 12 EU countries (plus a separate listing for West Germany and East Germany). The gross sample is the number of addresses drawn according to the sampling frame. The net sample is the number of successfully completed interviews. The completion rate (in percentages) is the net sample's share of the gross sample. The table shows that the completion rate of the EB41.0 varies

2 Country abbreviations are: B = Belgium; DK = Denmark; W-G = West Germany; E-G = East Germany; GR = Greece; ESP = Spain; F = France; IRL = Ireland; NIRL = Northern Ireland; I = Italy; Lux = Luxembourg; NL = The Netherlands; P = Portugal; GB = Great Britain; EU12 = European Union (B, DK, W-G, GR, ESP, F, IRL, I, Lux, NL, P, GB)

between 37% and 51% which is not different from the results obtained normally in the Eurobarometer studies. For more detailed information about the procedure the methodological reports of INRA in the written Eurobarometer reports can be consulted.

2.1.2 The FORSA telephone survey

From April 28 until June 3, 1994, 500 computer assisted telephone interviews (CATI) were carried out in each country of the EU which were administered centrally from the FORSA offices in Berlin and Dortmund. CATI guides the interviewer step by step through the questionnaire and stores keyed-in responses such that they are ready for immediate computer analysis.

The sampling basis for the study were paper telephone directories. No random dialling technique was used. Each country was divided into a number of regions, and the telephone numbers in each region were counted. The number of telephone addresses to be drawn from each region/telephone directory was computed in proportion to the number of inhabitants in the region. In this way an effort was made to correct for the differences in telephone ownership in the different areas. The columns or pages in the telephone directories were selected in a systematic way with a frequency calculated by the above mentioned procedure.

Final telephone addresses of potential respondents were drawn in a random fashion. Addresses which were clearly identified to be corporate lines were replaced with another random telephone number on the same column/page. The entire random choice procedure was computerised. Obviously, not every telephone address chosen that way belongs to a private household. The non-private households are removed from the list and substituted.

Training and supervision of interviewers is obviously facilitated by a centralised telephone interviewing operation as the one reported on here. After an oral briefing on the questionnaire the FORSA interviewers had to conduct at least three test interviews. One supervisor per (approximately) ten interviewers was present through the entire interviewing period. The specific FORSA-CATI-System of computerised dialling ensured that call-backs were made at the agreed-upon time, that busy numbers were re-dialled after a predetermined delay and that 'no answer' or 'not at home' coded addresses were re-dialled after a longer delay.

Because of the difference in alphabet the interviews in Greece were not done by CATI but by paper and pencil. Therefore it was not possible to use the usual nonresponse analysis. The number of recalls was fixed on 12 while the respondents who refused co-operation were not contacted any more. Telephone numbers which did not lead to a contact have been substituted by a random chosen number. Given this procedure for the field work the completion rate presented in table 2.2 was obtained.

Table 2.2 Completion rate FORSA feasibility study

	B	DK	G-W	G-E	ESP	F	IRL
Gross sample	1209	1001	1023	1013	1153	1254	959
Net sample	500	500	500	500	500	501	500
Completion rate (%)	41	50	49	49	43	40	52

	NIRL	I	Lux	NL	P	GB	EC without GR
Gross sample	372	1088	1388	1137	1124	1456	14177
Net sample	150	501	500	500	500	500	6152
Completion rate (%)	41	46	36	44	45	34	43.4

Details about the FORSA fieldwork can be found in a special report (FORSA, 1994). When tables 2.1 and 2.2 are compared, one can see that the completion rates of the INRA face to face study and the FORSA telephone study are at about the same level.

2.2.3 The telephone panel: A continuation of the EB41.0

The telephone panel study was based on the face to face interviews, such that only respondents who had participated in the face to face EB41.0 study could be selected for the telephone panel. The fieldwork was done from April 5 to April 30, 1994. It was centralised in Brussels, and all interviews in the three countries of Belgium, France and Spain were conducted from a central call-centre. The survey was run by the Company MARKETING UNIT - the Belgian INRA member Company- using the BELLVIEW CATI (Computer Assisted Telephone Interviewing) software.

The respondents were told that due to some technical problems their answers from the face to face interview had been lost and thus were kindly asked to answer some of the questions one more time. In the face to face interviewing phase an effort was made to secure the telephone address of respondents in Belgium, France and Spain. Not every interviewee had a telephone or was willing to reveal his or her telephone number, and therefore it was not possible to approach everyone of the 3124 Belgian, French and Spanish EB41.0 respondents for a re-interview. On the whole, 2352 first-wave respondents could be approached once more.

The Brussels agency conducting the interviews was also in charge of interviewer supervision and training. Call-backs were limited to a maximum of eight attempts. Four questions helped to screen the interviewees in order to make sure that identical respondents were re-interviewed: age, sex, occupation and subjective social class. Respondents who refused to co-operate were not contacted again.

Using this approach for the fieldwork, 884 respondents could be successfully re-interviewed. The following table gives the country-specific information on the completion rates.

Table 2.3 Completion rate EB41.Panel

EB41.Panel	B	ESP	F	Total
Gross sample EB41.0 = (respondents whose telephone number was known)	767	731	854	2352
Net sample (respondents who were sucessfully re-interviewed)	234	309	341	884
Completion rate (%)	31	42	40	37.6

It is clear from this table that the willingness to co-operate in the panel study was considerably less than in the original study. Also, one has to keep in mind that table 2.3 represents an additional selection stage since the face to face study already represents a selection from the population studied.

Details about this experiment can be found in a special fieldwork report (ZEUS, 1994). Table 2.4. summarises the different sampling methods used in the three studies. More detailed information from the national institutes co-operating in the INRA chain would have been desirable with regard to, for example, interviewer training and supervision, but this information was not available.

Although the overall completion rates are not so different, the table clearly points to differences in the procedures. Apparently given these differences, one can also expect considerable differences with respect to the background variables describing the populations in the various countries. Normally, such differences are corrected using unequal weights for the different respondents. Therefore, in the next section the weights specified in these studies will be discussed.

2.3 The weights of the data sets

The weighting variable corrects the national samples in such a way that the samples are brought as closely as possible in accordance with known distributions of the national populations with respect to socio-demographic characteristics. For each of the two studies such weights were estimated. Below the weights estimated by INRA for the face to face studies and by FORSA for the telephone studies are presented. The weights for the panel study need not be addressed because the weighting is less relevant.

2.3.1 Weights of the INRA face to face study

INRA generates the weights on the basis of target tables with the joint distributions of age by sex, and the distribution of the population with respect to region (Eurostat NUTS II), occupation, size of locality and size of household. The population data have been taken from Eurostat for all countries.Table 2.5 displays the means, minimum and maximum as well as the standard deviation of the weighting variable WSAMPLE used in the Eurobarometer EB41.0.

Table 2.4 Summary of the sampling methods

	EB41.0	**EB41.Panel**	**FORSA**
Type	face to face	telephone	telephone
Fieldwork	April 4th - May 6th	April 5th - April 30th	April 28th - June 3rd
Countries	all 12 EU member states	France, Belgium, Spain	all 12 EU member states
Completion rate	EU: 44,7%	37,6% (% of eligible households.)	EU:43,4%
Sample frame	- Census enumeration units (or otherwise)	- respondents of EB41.0 with a telephone who have given their number	Telephone directories
Selection method	- more than 100 sampling units per country are randomly chosen as start address - a random increment provides up to 10 addresses - one person/per household selected by next birthday or Kish method or an other	- all possible respondents are contacted - controlled by Age, Sex, Occupation and Subjective Social Class	- From 10 to 22 'provinces' per country samples are drawn according to the size of the province's population - one person/per household selected by next birthday method
Interviewers' testing and supervision	INRA's national associates are responsible	INRA central - computerised dialling	FORSA central - tests in advance - computerised dialling
Call backs	2 revisits	8 call-backs	12 call-backs
Refusals	no refusal reversion	no refusal reversion	no refusal reversion
Substitution	random walk	no substitution	by random number

Table 2.5 The weights estimated by INRA

WSAMPLE	**Mean**	**Minimum**	**Maximum**	**Std. Dev.**
France	.97	.43	2.06	.26
Belgium	.92	.35	2.62	.30
The Netherlands	1.00	.12	4.93	.53
West Germany	.97	.26	3.54	.47
Italy	.95	.12	3.06	.39
Luxemburg	1.00	.35	3.25	.47
Denmark	1.00	.35	2.42	.35
Ireland	1.00	.45	3.70	.33
Great Britain	.95	.24	2.49	.26
Northern Ireland	.98	.88	1.10	.06
Greece	1.00	.38	2.33	.21
Spain	1.00	.19	1.97	.32
Portugal	1.00	.41	2.85	.32
East Germany	.95	.36	2.14	.29

2.3.2 Weights of the FORSA telephone survey

FORSA has corrected for the differences of the sample size and population size in each region by providing different inclusion probabilities to respondents of different regions. Furthermore FORSA constructs weights to approximate the joint distribution of age by sex. Table 2.6 displays means, minimum and maximum as well as standard deviation of the weighting variable WFSAMPLE used in the FORSA survey:

Table 2.6 The weights estimated by FORSA

WFSAMPLE	Mean	Minimum	Maximum	Std. Dev.
France	1.00	.68	2.35	.35
Belgium	1.00	1.00	1.00	.00
The Netherlands	1.00	.62	4.10	.46
West Germany	1.00	.71	1.49	.19
Italy	1.00	.47	1.81	.36
Luxemburg	1.00	.70	2.84	.34
Denmark	1.00	.72	2.34	.33
Ireland	1.00	.73	1.78	.27
Great Britain	1.00	.70	2.21	.33
Northern Ireland	1.01	.70	2.21	.34
Greece	1.00	.64	2.39	.38
Spain	1.00	.52	2.99	.40
Portugal	1.00	.63	2.28	.37
East Germany	1.00	.57	3.37	.35

The difference between these two sets of weights will be due to the different target table and the differences between the samples. Whereas the FORSA weights approximate a simple age-by-sex distribution, INRA specifies a more complex target table that incorporates additional variables. Besides that INRA uses as population statistics Eurostat data while FORSA has used the statistical information from the statistical offices of the different countries.

Due to this difference in approach different characteristics of the national samples can be expected. In order to make the samples comparable, one should use the same population figures and the same variables. This will be done in the next chapter where also the results for the substantive variables after weighing will be compared.

2.4 Comparison of item nonresponse rates

In this last section the face to face study and the telephone study will be compared with respect to item nonresponse because it presents another indication of the differences in the field work of the different organisations. A number of often used closed-ended questions have been selected to investigate possible differential nonresponse rates. In table 2.7, the combination of 'don't know', 'no answer' and 'refused to answer' codes between the face to face (INRA) study and the telephone (FORSA) study are compared.

With respect to item nonresponse no remarkable differences could be observed between the two studies on European level. There are questions that produce more refusals than others, but those differences appear both in the INRA face to face and the FORSA telephone survey.

Table 2.7 The percentage of 'don't know', 'no answer' and 'refused to answer' codes in the face to face EB41.0 and telephone (FORSA) study

Variable	EB41.0	FORSA
Satisfaction with life	0.6	0.5
Satisfaction with democracy	3.9	4.8
Frequency of political discussion	0.7	0.5
Persuade friends to share opinion	1.7	1.5
Watching news on television	0.2	0.0
Reading news in daily papers	0.3	0.1
Listen news on the radio	0.3	0.1
Interest in European politics	1.1	0.9
Informed about European politics	1.9	2.3
Is membership in EU good-bad thing	5.0	6.8
Has country benefited from EC membership	17.0	18.6
Will R vote in EP election	6.0	3.9

National differences were also studied but are not reported here. This detailed analysis shows that there is national variation in certain questions: some nations display more refusals with regard to these questions than others. These national differences, however, also appear in both surveys. As a result, it can be stated that there are no systematic differences between the institutes (FORSA vs. INRA) with respect to item nonresponse.

2.5 Conclusion

It was the purpose of this chapter to discuss the differences in sampling design, fieldwork and to look at the consequences for the total non response, the necessary weights and the item non response.

The findings indicate that:
- the total nonresponse looked very similar,
- the item nonresponse was also quite similar, and
- the weights differed substantially.

The first two findings do not imply that the samples are equivalent; it only indicates that the co-operation of the sampled respondents was similar. Due to the differences in the sample designs the originally drawn samples were already quite different. For example we expect that the ownership of a telephone could make a difference between households. This problem will be elaborated in chapter 4. Due to differences in the fieldwork, especially the number of recalls and the substitution of non co-operating households, the final samples certainly were even more different. For these reasons the necessary weights were also quite different. These points will be further evaluated in the next chapters.