

6. TECHNICAL ANNEX

A total of 5 141 women born between 1938 and 1967 were selected from the Central Population Register by simple random sampling in June 1989, and a total of 2 048 men born in 1943-47, 1953-57 and 1963-67 were selected in the same manner in June 1992. Thus the men were interviewed three years later than the women. The sample for the men was drawn so that the oldest cohort would be the largest, the middle cohort the second largest, and the youngest the smallest. The final sample sizes were 5 105 women and 2 040 men after excluding those who had died, had moved abroad, or were living in institutions.

Sample selection

A pilot study was performed from March to April 1989, before implementation of the actual survey, with the aim of testing the feasibility of the questionnaire and its items with respect to achievement of the objectives. The test interviews, ten of which were tape recorded, yielded information on the expected approximate duration of the actual interviews and the costs involved. Interviewers underwent the necessary training before the pilot study.

Data collection

The test sample contained 164 women and 96 men, and the data were collected by personal interview, as in the survey proper. After the pilot study had shown that the financial resources available would be insufficient for interviewing both women and men in 1989, the decision was made to study only women at that stage. It was only in Spring 1992 that it became possible to interview men, and this was done in the autumn of that year.

The women were mostly interviewed between August and December 1989, with some interviews conducted during the first three weeks of January 1990. The work was done by female interviewers from Statistics Finland who had attended a one-day training course for that purpose in August 1989. In addition to the questionnaire, there was also a separate 'time chart' for the collection of data on the respondent's life history. The purpose of the chart was to help the interviewee to remember when given events had taken place in her life and to allow the interviewer to assess their internal consistency.

The men were interviewed between September and December 1992, the great majority of them by female interviewers and a small minority by male interviewers from Statistics Finland who had attended a one-day training course in August or September 1992. No 'time chart' was used in this case.

The mean duration of the interview was 1 hour and 12 minutes for the women and 1 hour and 4 minutes for the men. The men's questionnaire was much shorter.

The data processing procedures were the same for both samples. Codes indicating occupation, local government district and education were inserted on the forms at Statistics Finland, after which the forms were read through. The checked forms were then transferred to a computer and finally computer checked using special programs.

Sample bias and non-response

Non-response was 18.6 per cent for the women and 18.1 per cent for the men. No adjustment for non-response was made to the data on the women. The analysis of non-response in the women's data is somewhat restricted because marital status, for example, had been omitted. This was added later, but at such a late stage that the information could not be used for adjusting for non-response. There are strikingly large percentages of non-response in some groups (table 6.1, d).

Table 6.1
The survey population and non response

Age group	Women			Men		
	Single	Married	Previously married	Single	Married	Previously married
a. Number of eligible persons according to national statistics						
22-24	83071	23115	1312			
25-29	88721	88411	7650	131825	54925	4279
30-34	49140	124763	15056			
35-39	32842	143533	23874	60236	124051	19314
40-44	24078	155359	33137			
45-49	14312	106748	26257	28184	137437	28451
50-51	5436	42400	11466			
b. Number of persons in target sample						
22-24	369	103	4			
25-29	426	395	42	383	135	9
30-34	221	616	72			
35-39	142	683	98	186	373	59
40-44	118	706	165			
45-49	57	488	131	126	649	120
50-51	20	192	57			
c. Number of persons interviewed						
22-24	319	88	4			
25-29	331	358	31	315	121	6
30-34	173	540	58			
35-39	90	569	77	131	323	45
40-44	87	570	119			
45-49	37	380	107	88	552	89
50-51	16	157	44			
d. Percentage of target sample not interviewed ^{a,b}						
22-24	13.6	14.6	..			
25-29	22.3	9.4	26.2	17.8	10.4	33.3
30-34	21.7	12.3	19.4			
35-39	36.6	16.7	21.4	29.6	13.4	29.6
40-44	26.3	19.3	27.9			
45-49	35.1	22.1	18.3	30.2	14.9	25.8
50-51	20.0	18.2	22.8			

^a Computed as $100 * (b - c)/b$.

^b Marital status for women, who were not interviewed, is their marital status at end-year 1989, and not at the date the sample was drawn. This means that in some cases it can be different from their marital status at interview date.

Furthermore, the women's data showed fairly large differences in willingness to respond between the provinces when divided into cohorts, and this should be taken into account when using variables that are dependent on region and age.

A more thorough study was made of non-response among the men. When the sample was selected from the registers, both marital status and place of residence were included as background information, and information on the level of education at the end of 1991 was added later while data on biological children was only added at the end of 1992. When the analysis of non-response was performed at the end of 1992, the information on the level of education was not yet available.

It was more difficult to interview men who had been previously married, and also single men born between 1943 and 1957, than men who were married at the time of the survey. As expected, the non-response rate was higher for men without children, being 25 per cent for those born between 1943 and 1947 and 29 per cent for those born between 1953 and 1957. Non-response also varied according to the level of education, in that the less education a male respondent had, the more likely it was that he would decline to be interviewed. The non-response rates for men with a basic general education varied between 21 and 24 per cent in the 5-year age brackets. Non-response also varied by region, being highest for men born between 1953 and 1957 and living in the province of Uusimaa, 30 per cent, whereas it was generally well below 20 per cent in almost all of the other regions of Finland.

The bias in the data on men caused by non-response was estimated in the following manner. The base population can be divided into two strata according to whether or not the person could be interviewed. Let us say the sizes of these strata are N_1 and N_2 and the corresponding stratum weights are $W_1 = N_1/N$ and $W_2 = N_2/N$, where subscript 1 refers to those who were interviewed and subscript 2 to the non-respondents. The data obtained represent a sample from stratum 1. Let us call the sample average in the respondents stratum y_1 . If the averages Y_1 and Y_2 in the basic population differ, the sample average is a biased estimator for the base population average and the magnitude of the bias is:

$$\text{Bias}(\bar{y}_1) = E(\bar{y}_1) - Y = W_2(Y_1 - Y_2)$$

Since our calculations deal with percentages they are computed by:

$$\text{Bias}(p_1) = w_2(p_1 - p_2),$$

which is an application of the above equation to averages. Since exceptionally large biases were observed in all the variables studied, it was decided that major errors would be corrected by post-stratification.

The post-stratification was carried out by cohort, using marital status and region as variables. The region variable comprised two categories, i.e. the metropolitan area of Helsinki vs. the rest of Finland, and the marital status variable was grouped into three cate-

Post-stratification

gories, with divorced and widowed grouped together. Thus six strata were obtained for each cohort.

Conditional and unconditional inference are related to the theoretical aspects of post-stratification. According to conditional inference the sample sizes are known after selection and can be used as absolute figures. The mean and standard error can be calculated using stratified random sampling.

$$Y_{str} = \sum N_h * \frac{y_h}{n_h},$$

the weighting coefficient w_h is thus N_h/n_h , where N_h is the number of men in the base population and n_h the number of men in the sample in stratum h. The calculations were computed for all three 5-year age brackets.