6. TECHNICAL ANNEX

SURVEY DESIGN

The 1994 Slovenian Family and Fertility Survey (FFS) was designed in accordance with the international standards set for the FFS. These were based on the recommendations of the international expert group that guided European Family and Fertility Survey project. The implementation was adapted to previous Slovenian experience in population sampling, and the sample size and structure were discussed with leading sampling experts.

The initial plan was to include 3000 women and 2000 men of reproductive age. Regardless of the overall size of the population, sample size is the key factor determining the extent to which small subgroups can be analysed; the larger the sample, the more detailed the analyses that are possible. A sample of 3,000 women in the Slovenian FFS permitted at least rough analysis of the data for groups as small as one per cent of the whole female population of reproductive age. To analyse smaller groups, a significantly larger sample would have been needed.

The sample was implicitly stratified by 12 regions, and a three-stage design was used. The primary sampling unit (PSU) was the community. Out of a total of 1,500 PSUs, 300 were selected with a probability proportional to the number of women aged 18-45. The secondary sampling unit (SSU) was the enumeration area, of which there were 12,000 in the entire country. Three SSUs were selected within each PSU, again with probability proportionate to the number of women aged 18-45. In two of the three SSUs, only women were to be interviewed; only men were to be interviewed in the remaining one.

Since the target numbers of respondents were 3,000 women and 2,000 men, in theory it was necessary to interview either five women or seven men per SSU according to the sex designated for that SSU. Where there were not enough potential respondents in a given SSU, every eligible individual was selected, i.e. all women or men aged 15 - 45.

Of course, in anticipation of non-response, the actual number of persons selected had to be larger. Based on experience in earlier Slovenian surveys, it was estimated that nine women or 14 men, respectively, would be sufficient to allow for non-response. In the first wave of the fieldwork, however, the interviewers were given just seven addresses of women or 10 addresses of men, respectively, per SSU. Assuming that the level of non-response would turn out to be relatively low, these numbers were expected to achieve the required numbers of respondents. In addition, the addresses of two women or four men, respectively, were assigned to a supplementary list to be called upon should it become obvious that the initial lists were not going to provide enough respondents to reach the target numbers. After a few months, this proved indeed to be the case, and almost all the supplementary addresses in Ljubljana along with about half of those outside Ljubljana were used.

Use of the supplemental address list should be distinguished from the substitution procedure. Substitutions were made immediately when an interviewer decided that an individual who had been selected could not be interviewed. The supplementary list, however, contained additional addresses selected by the project management to be used when it was judged necessary in order to achieve the desired sample size. The latter avoided possible biases inherent in the substitution procedure, e.g. interviewers using substitutions to avoid the effort of reaching individuals who were more difficult to contact.

The final selection of individuals was taken from the Central Register of the Population of the Republic of Slovenia. Due to confidentiality restrictions, however, only the addresses of the persons selected could be used and not their names. This resulted in problems in buildings containing multiple dwelling units, where the number of persons targeted had to be increased. Since the population register does not identify households, the only source of the necessary data was the 1991 census. The number of persons selected in each multi-unit building was enlarged according to the ratio of the total number of households to the number of households with at least one eligible individual, using census data on the number of households at a given address and the number of households containing target persons at that address.

INTERVIEWING

The interviews were conducted between December 1994 and December 1995. This long period had certain methodological consequences for the computation of the respondents' age, especially with respect to the youngest age group.

Because the interviewers had only the addresses of potential respondents and not their names, additional dwelling units had to be selected in multi-unit buildings (approximately one third of the sample). Randomisation tables were used for this purpose. This selection was carried out separately before the interviewing stage by enumerators specially trained for the task. In single-unit buildings, the address itself identified the eligible dwelling.

Every address selected was screened by the interviewer who made a list of all persons in every household. The eligible person with the most recent birthday was then identified as the target individual. This screening procedure excluded all households at addresses where no eligible individuals were found. Out of a total of 11, 473 addresses on the list, including the 2,703 from the supplementary list, 10,434, or 91 per cent, were actually used in the survey.

The two procedures described above naturally reduced the completion rate. A total of 4,559 persons were interviewed. While the response rate does not seem unusual for this kind of survey, the completion rate (interviews performed as a fraction of addresses contacted) was only 43 per cent, which is relatively low. There are a number of reasons for the dropout:

- wrong address, failure to locate a dwelling (5 per cent of all addresses);
- failure to screen a selected dwelling (12 per cent of all dwellings located);
- screening showed the household contained no eligible person (40 per cent of all contacted households);
- non-response from selected eligible persons (14 per cent of all selected persons).

The high percentage of households with no eligible person is somewhat surprising and remains partly unexplained. A certain percentage of households without an eligible person was expected because of the need to increase the number of households in multiple-dwelling buildings as discussed above. In other words, it was inherent in the sample design that there would be a certain number of dwellings without an eligible person. In addition, the problem may be accounted for partly by the population register being out of date and partly by hidden refusals to participate in the survey. It is also possible that there were some inconsistencies in the recording procedure or in compliance with the interviewing instructions.

When non-response is confined to explicit refusals, non-contacts and other minor categories, the rate is 14 per cent, which is fairly good. Non-response due to failure to reach an eligible individual was low since the interviewers were instructed to try to contact each person five times. In general, non-response includes all individuals who were eligible for the survey and were eventually located but dropped out during the data collection stage, usually because of absence or refusal but also sometimes because of inability to participate or other reasons. The rate of refusal (refusal to participate by a person who was contacted) was 10 per cent, which is very good for this kind of survey.

The low completion rate also includes some hidden categories of non-response, but the main explanation lies elsewhere. A principal cause was failure at the screening stage, i.e. when a household did not even permit identification of an eligible person or when it was not possible to gain access to a household or households in a selected dwelling (12 per cent of all dwellings located). Assuming that the proportion of eligible households and the rate of non-response would be the same in dwellings (households) where screening failed as in screened dwellings, the total rate of non-response was 25 per cent. It is quite possible ineligible households were actually more common in households that were not screened (e.g. the dwellings were more likely to be uninhabited), but there may also have been hidden non-response among ineligible persons (during household screening, a certain number of non-response cases could have been entered as ineligible persons), which makes 25 per cent a reasonable estimate. Similar adjustment of the refusal rate raises it from 10 per cent to the actual 16 per cent.

Thus the level of non-response is high enough to be a matter of concern. Comparable levels can be found in FFS surveys in other European countries, of course, but this does not change the fact that data do not exist for a quarter of the eligible population. If this quarter does not differ significantly from the interviewed population, non-response does not pose any particular problem. If that is not the case, however, generalisation of the results to the total population becomes problematic. Whether those who were missed differ from the others with respect to key variables necessarily remains unknown, and this is the core of the non-response issue, which may well deserve more detailed analysis.

WEIGHTING

Analysis showed that the survey data roughly matched the target population with respect to critical variables. Naturally, the usual measurement error found in surveys of this type has to be taken into account, e.g. underestimation of certain phenomena, such as abortion. Weights needed to be introduced, however, to correct for the most basic causes of deviation in the sample:

- SSUs that were too small and did not have enough people;
- the varying non-response rates of PSUs and SSUs;
- the discrepancy in age range between that used to calculate probability proportional to size in the selection of PSUs and SSUs (ages 18-45), and the selection of eligible individuals

(ages 15-45);
- the elevated likelihood that any given person would be selected in a household with several eligible individuals living in a multiple-dwelling building;
- the differences among SSUs in the number of addresses used;
- the failure to match the basic characteristics of the population at large (sex, age, region).

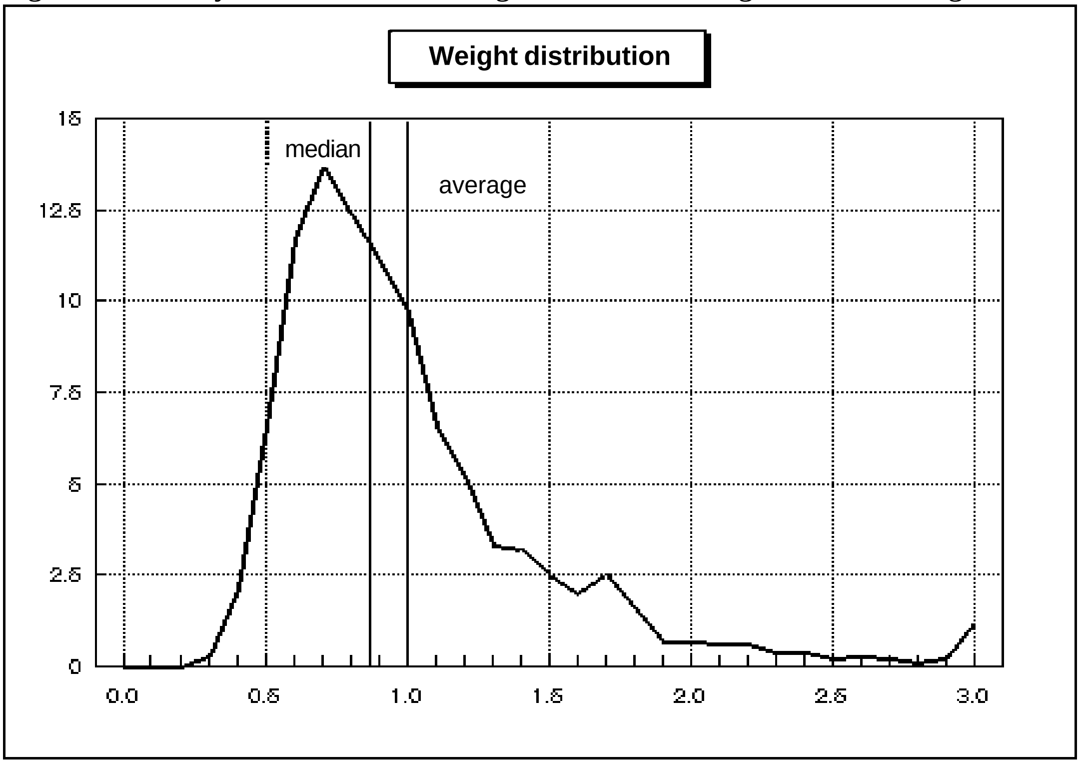
The basic weighting procedure adjusted the sample for varying non-response rates in SSUs and for the inconsistency in the age range used to select PSUs and SSUs, on the one hand, and eligible persons, on the other. The weighting was based on the number of eligible persons in the population register, making special weighting for the difference in the age range unnecessary.

In the second stage, an adjustment was made, based on register data, for three control variables: sex, age (5-year age groups) and size of settlement (under 500 inhabitants, 500-1,999 inhabitants, 2,000-9,999 inhabitants, 10,000-99,999 inhabitants, and 100,000 or more inhabitants, i.e. Ljubljana and Maribor). Hence, if the survey data showed a specific age/sex/settlement cell with 80 respondents where the population register indicated that there should have been 90 respondents, a weight of 90/80 was applied to the data for this cell. It should be pointed out that the age variable was based on "age at last birthday" as of the interview date and thus depended not only on year of birth but also on the date of interview.

When the weighting was completed, the usual practice was followed to avoid extreme values: the weights were truncated at 0.3 and 3 and scaled to the total number of 4,559 respondents. The average weight then becomes equal to 1.0. The coefficient of variation for these weights is calculated as: cv(w)=(standard deviation of the weights)/(mean weight); the increase in variance due to the weighting can be expressed as ($1 + cv^2(w) = 1.22$) or 22 per cent and needs to be taken into consideration.

Figure 6.1 shows the final distribution of the weights; the skewing of the distribution toward the left is very obvious. It is due to under-representation of the 15-19-year age group. Because

Figure 6.1: Density distribution of the weights, the median weight and mean weight



women in this age group made up only 9.6 per cent of the sample as opposed to 15.8 per cent in the general population, they received a relatively large weight. The discrepancy can be accounted for to a large extent by the selection procedure within households where both mother and daughter were eligible. Another reason is the year-long duration of the interviewing, which meant that many of the 15-year-olds selected were no longer age 15 by the time they were interviewed. In an unweighted analysis, deficient numbers in the youngest cohort would have had a significant influence on certain variables that are closely associated with age. The following are examples of such variables:

- the proportion of women who were economically active, which falls from 72.1 per cent in the unweighted sample to 67.4 per cent in the weighted sample because so many young women were studying;
- the proportion of women who had ever been married, which drops from 7.1 per cent in the unweighted sample to 64.4 per cent in the weighted sample;
- the proportion of women who had had at least one live birth which declines from 76.2 per cent in the unweighted sample to 69.5 per cent in the weighted sample.

The weighting has absolutely no impact on the analysis within individual age groups, however. And since analysis by age group is essential with data of this type, the weights described above are not actually so radical.

In general, weights have a beneficial effect since the weighted distributions for most demographic variables (education, marital status, economic activity) come close to population estimates and the results of other surveys, e.g. the 1995 Labour Force Survey (1995 LFS). Tables 6.1-6.4 compare the unweighted and weighted sample distributions and population estimates for a number of variables.

Table 6.1: Sex and age

	MEN		WOMEN	
	Unweighted	Weighted	Unweighted	Weighted
Total	100.0	100.0	100.0	100.0
15-19 years	15.3	16.7	9.6	15.8
20-24 years	12.1	15.5	13.1	15.5
25-29 years	14.5	16.0	18.6	16.7
30-34 years	17.8	16.5	20.3	17.1
35-39 years	18.3	17.3	19.7	17.4
40-45 years	21.9	18.0	18.8	17.6

Table 6.2: Marital status

	MEN			WOMEN		
	Unweighted	Weighted	1995LFS	Unweighed	Weighted	1995LFS
Total	100.0	100.0	100.0	100.0	100.0	100.0
Single	39.9	44.0	52.0	28.3	35.6	39.6
Married	57.0	53.5	42.1	67.1	60.6	51.2
Widowed	0.3	0.2	0.2	0.7	0.5	1.0
Divorced	0.9	0.8	1.1	0.9	0.7	3.1
Separated	1.9	1.4	-	3.0	2.6	-
Extramarital	-	-	4.7	-	-	5.1

Table 6.3: Level of education

	MEN			WOMEN		
	Unweighted	Weighted	1995LFS	Unweighted	Weighted	1995LFS
Total	100.0	100.0	100.0	100.0	100.0	100.0
Unknown	0.6	0.5	-	0.4	0.3	-
Incomplete	1.4	1.4	6.3	2.4	2.0	4.7
primary school	23.0	23.4	25.3	27.5	29.7	28.9
Vocational	37.1	37.4	33.1	26.1	24.5	22.2
Secondary	26.5	27.2	25.6	29.9	30.8	31.2
College degree	5.5	4.9	4.5	8.1	7.4	7.8
University	5.0	4.4	4.5	5.4	4.9	4.7
MSc/MA, PhD	1.0	0.8	0.6	0.5	0.4	0.3

Table 6.4: Type of settlement

	MEN			WOMEN		
	Unweighted	Weighted	Population	Unweighted	Weighted	Population
Total	100.0	100.0	100.0	100.0	100.0	100.0
Unknown	0.4	0.4	0.2	0.3	0.3	0.1
Urban	52.1	48.7	49.4	50.1	52.0	52.3
Rural	47.5	50.9	50.4	49.5	47.7	47.5

As evident from Tables 6.1 - 6.4, deviations from the population values are small and acceptable. Similar results were obtained for some other variables, e.g. the regional structure. No further corrections for type of settlement, region, marital status, education, etc., were made. Because the underestimation of single persons was relatively large, however, especially that of single men, and the characteristics of single persons vary considerably by age, it would have been appropriate to include this variable in the correction process.

PRECISION OF THE RESULTS

The survey is based on a complex, three-stage cluster sample. In addition, a relatively broad range of weights was applied. All this leads to a considerable widening of confidence intervals. In this section, the precision of the results is estimated for some target variables. The sample variance was computed at the *Statistical Office of the Republic of Slovenia* using special software (SUDAAN).

The precision of estimates is usually expressed by the design effect. This is the ratio between the precision of a given sample, expressed in terms of variance, and the precision of a simple random sample of the same size. A simple random sample thus has a design effect of 1.0. The confidence intervals are widened by a factor equal to the square root of the design effect. In Tables 6.5 - 6.11, the design effect (*deff*) is around 1.8 for the usual variables expressed as shares of the interval (0.1, 0.9), or 10-90 per cent of the interval. The results agree with those obtained in other surveys (e.g. the Slovenian Public Opinion Survey). In analysing the results, special attention should be given to variables that are particularly dependent on local circumstances, e.g. items related to religion or social status. The design effect for such variables may be significantly higher.

The effect of the above may be illustrated as follows: the confidence intervals for the variables (shares) are widened by approximately 20 to 40 per cent. The estimate is valid only for shares between 10 and 90 per cent. When the share values are extreme, the effect is significantly smaller but, for certain variables, it can be significantly greater.

The tables also provide the coefficient of variation (CV = standard error / share) for the variables under discussion. It is evident that the estimates of shares are relatively precise, as CV is almost always below 5 per cent.

The estimates in these tables are based on the total sample, of course. Estimates for subgroups are less precise, depending again on the size of the subgroup. In borderline cases, the following rule of thumb can be applied: when there are fewer than 10 observations, the estimates should not be published at all because they are too unreliable; when there are 10-30 observations, the data should be considered as extremely imprecise; and when there are 30-80 observations, the estimates should be presented with some caution and reservations. Only when there are at least 80 observations can the CV be expected to reache the level of 10 per cent or less.

Table 6.5: Total number of children

Average	*St.error*	*CV*	*deff*	*Confidence interval*	
1.19	0.02	1.68	1.73	1.15	1.23

Table 6.6: Ever married

	Percentage	*St. error*	*CV*	*deff*	*Confidence interval*	
Yes	61.02	0.97	1.59	1.81	59.12	62.92
No	38.98	0.97	2.49	1.81	37.08	40.88

Table 6.7: Ever had a live birth

	Percentage	*St. error*	*CV*	*deff*	*Confidence interval*	
Yes	65.05	0.94	1.45	1.79	63.21	66.89
No	34.95	0.94	2.69	1.79	33.11	36.79

Table 6.8: Partner’s drinking problem is a sufficient reason for divorce

	Percentage	*St. error*	*CV*	*deff*	*Confidence interval*	
Yes	70.31	1.02	1.45	2.16	68.31	72.31
No	29.69	1.02	3.44	2.16	27.69	31.69

Table 6.9: Ever had an abortion for maternal health reasons

	Percentage	*St. error*	*CV*	*deff*	*Confidence interval*	
Yes	95.55	0.37	0.39	1.42	94.82	96.28

No	4.45	0.37	8.31	1.42	3.72	5.18

Table 6.10: Ever had an abortion because the foetus was not healthy

	Percentage	*St. error*	*CV*	*deff*	*Confidence interval*	
Yes	91.34	0.56	0.61	1.74	90.24	92.44
No	8.66	0.56	6.47	1.74	7.56	9.76

Table 6.11: Are you religious?

	Percentage	*St. error*	*CV*	*deff*	*Confidence interval*	
Yes	48.18	1.30	2.70	3.00	45.63	50.73
To an extent	26.49	1.01	3.81	2.33	24.51	28.47
No	25.33	1.17	4.62	3.22	23.04	27.62

Of course, when it comes to other variables, the conclusions drawn from the above tables can serve only as a general guide. It is particularly important to keep the issue of precision m in mind when making comparisons among subgroups or years. Statistical packages usually assume a simple random sample, and less precision typically results in lower correlations. When greater precision is desirable for particularly important variables, analysis of each specific issue is recommended.

CONCLUSIONS

In sum, given the substantial constraints (to protect data at the individual level), the sample design was optimal and took into account all available resources (census, register of population). Similarly, at the weighting stage, all available information was used to match the sample as closely as possible to the socio-demographic structure of the population. The remaining deviations fell within the range common in such surveys.

The topics covered in the survey were extremely sensitive. If we add to that the practical obstacles associated with the lack of information specifically identifying the individuals selected for the sample, the actual non-response rate of about 25 per cent is reasonable. Similarly high non-response rates are not uncommon in developed countries. Lower rates can only be achieved with a more adequate interviewing strategy, which significantly increases the survey cost.

For most analyses, it seemed appropriate to use weighted data even though the weighted and unweighted results were not expected to differ very much. Women aged 15-19 are the exception, since they are correctly represented in the total sample only after weighting. Hence the use of weights was indispensable for analyses based on the total sample rather than age groups. For simple analyses (means, proportions, distributions) within age and sex subgroups, weights are recommended, but they do not affect the results very much. In more complex statistical analyses (regression, analysis of variance, etc.), a certain amount of caution is appropriate since some standard statistical packages are not adjusted for the use of weights. Moreover, the effect of weighting is somewhat smaller in such analyses and may be conceptually controversial.

If confidence intervals are computed, it should be borne in mind that the usual formula for shares, var(p)=pq/n, underestimates the confidence interval by a factor equal to *deff*. As specified in the section on *Precision of the Estimates* above, the *deff* values for shares in the FFS range from *deff*=1.42 for *maternal health-related abortion* to *deff*=1.79 for *ever had a live birth* to *deff*=3.00 for *religiosity.* The *deff* factor can affect the standard *t-test*, which is normally used to compare differences between two groups; if high *deff* values are ignored, the *t-test* values may be too high and statistical significance may be ascribed to differences that do not actually warrant it.