Appendix 5 Surveys used

For *Children and their families 2002*, four sample surveys have been used: Statistics Sweden, Living Conditions Surveys (LCS) 2001 and 2002; Statistics Sweden, Labour Force Surveys (LFS) 2002; Statistics Sweden, Income Distribution Survey (HEK) 2001 and the National Agency of Education, Parents Survey (FU) 2002.

In the Parents Survey, children are both a sample unit and a survey unit. Data are provided by the children's guardians. In all other surveys, the sample person and our survey person, children, are not the same. A sample of adults is taken who give data on the children that live in the household. This influences different probability distributions in primarily two respects:

The **first** is that the child has a heightened probability of being included in the sample. In LCS, the child is included through either the mother or the father being included in the sample. The probability for being included in studies therefore doubles for children that live with both parents. Different estimators must be corrected for this.

In LFS, however, it is possible to choose either "all mother's children" plus children of single fathers or vice versa. In each case, each child has the same sample probability as the mother or the father. There is no issue here of double sample probabilities.

The **second** is that estimates of proportions of some characteristic related to children are constantly ratio estimates. The base, i.e. the total number of children in the sample, is random due to a varying number of children in the families. The total number of children in the sample can be said to be the total of "clusters" of different sizes. When estimating variances for population totals, however, this does not affect the calculations. The adjustment coefficient in LCS can be used directly only if it is divided by two when children have cohabiting parents. The adjustment coefficient in HEK is already adapted to reporting households and the same coefficient can be used for statistics on children. In LFS, the same adjustment coefficient as for mother and father is used. In FU, the adjustment coefficient is from the beginning designed for reporting on children.

Labour Force Surveys (LFS)

The Labour Force Surveys are carried out every month. From the survey, monthly, quarterly and annual averages are calculated. The objective of the survey is to describe the current employment conditions and to give information on the development of the labour market.

Survey population

The survey population is all registered persons in Sweden between the ages of 16 and 65.

Sampling frame

Statistics Sweden's Total Population Register

Sample

The sample consists of three separate samples, one for each month in the quarter. Each sample, which for each month is made up of around 21 000 persons, is rotated in a way so that one eighth is exchanged between two for each following survey round. For each sample then, this occurs every three months. In other words, persons in the sample are interviewed once per quarter and, in total, eight times over a two-year period, after which time they are exchanged for new sample persons.

The sample is drawn at the end of the first quarter every year to cover the coming year's need for sample persons. When drawing the sample, a division into strata is carried out, by county, sex, citizenship (Swedish, foreign) and employment (employed, unemployed) according to the Employment Register kept by

Statistics Sweden. This leads to a total of 192 strata. Within each stratum, a systematic probability sample is drawn. The register is also sorted into age within the strata. A full year's sample is used for statistics on children.

Method of data collection

Telephone interviews are used. In some cases, for example, if sample persons do not have a telephone, a face-to-face interview is carried out.

Non-response

There was 15.9 per cent non-response in 2002. Extra information is used when there is non-response from Statistics Sweden's Employment Register, the Total Population Register and the Job Seekers Register at the Swedish Labour Market Board.

Statistics on children

When producing statistics on children, the study object consists of the children of sample persons. Sample persons can either be the mother or father. Interview persons state how many children there are in the family and when the children were born (year and month). The statistics are based on data from the mother or father.

Parents Survey (PS)

In 2002, the Parents Survey was carried out by Statistics Sweden on commission from the National Agency for Education. It is the follow-up to previous years' Childcare Survey but with partly changed content. A primary objective was to illustrate the different forms of childcare that children in different population groups take part in, if people are satisfied with the childcare forms and if not, what would be the alternatives.

Survey population

Families with children living at home, aged 1-5 (born 010197-310801), and children aged 6-12 (born 010190-311296).

Sampling frame

The sampling frame for the survey is Statistics Sweden's Total Population Register 30-06-2002.

Sample

From the group of younger children, a stratified sample of 71 935 children was drawn, the stratification variable being the municipality. The sample probability varied greatly between the municipalities. From the group of older children, an unstratified sample of 11 000 children was taken.

Method of data collection

Postal questionnaires were used, complemented by non-response follow-up by telephone and some collection from registers.

Non-response

The weighted non-response was 9.6 per cent in the younger group and 13.2 per cent in the older.

Statistics on children

The survey is from the beginning designed to give statistics on children.

The Household Economy (HEK)

The Survey on the Household Economy aims to illustrate income conditions for Sweden's entire population. Reporting is carried out regarding households as well as individuals. Household refers, in this case, to a housekeeping unit. A housekeeping unit consists of persons who live in the same home and have a joint economy and joint housekeeping. The presentation of the survey results uses the concept "full year household", which described households in which all adult individuals (aged 18 or older) were registered in the national population register both on January 1 and December 31 of the same year.

Survey population

All households with persons aged 18 or older.

Sampling frame

Statistics Sweden's Total Population Register

Sample

The sampling frame in 2001 included two strata, one for persons aged 18-74 and one for persons aged 75 and older. The sample was drawn as a

stratified sample with independent random sampling. The total sample in 2001 was around 15 200. During the period 1993-2001, the sample has varied in size between 13 000-19 000. In 2001, roughly 5 600 children aged 0-17 were included in the sample.

Method of data collection

Data collection is carried out partly directly by telephone interviews and partly using the registers of different authorities.

Non-response

Non-response was around 28 per cent in 2001. However, all the register data is available for non-response persons. Missing data on family structure are replaced by data from the population register.

Statistics on children

The original sample has been redefined to apply to a population of children. The children included in the sample for statistics on children are those that are included in the chosen family units, according to what is reported by the sample person.

Living Conditions Survey (LCS)

Statistics Sweden carries out current annual surveys on living conditions. The Living Conditions Survey is an annual sample survey of prosperity - both the present level and changes over time, the relationships between problems in different areas and the differences between different population categories. The survey contains a basic set of questions that recur every year. In addition, a number of in-depth areas are included, according to a rolling schedule, with each area recurring every eight years for a two-year period.

Survey population

The sample is taken from persons aged 16-84. Children aged 10-18 in the households of sample persons are included in the survey but they are asked different questions than those asked to the sample persons.

Sampling frame

Statistics Sweden's Total Population Register

Sample

Every year, a systematic sample of 7 000 - 8 000 persons is drawn according to age. From the sample, persons who have taken part in the survey at some point over the last seven years are removed. A part of the sample is made up of a panel that has taken part in the LCS during the previous round with the relevant in-depth area. The panel is supplemented with immigrants and persons that have just "grown into" the population (16-23 years).

Results are generally presented for a two-year survey, which has a sample of roughly 15 000 persons. Roughly 6 500 children aged 0-17 belong to the sample persons from one two-year survey.

Method of data collection

Primarily face-to-face interviews but, to a certain degree, telephone interviews. In addition there is comprehensive collection from registers.

Non-response

Non-response was 25 per cent in 2002. When estimates are produced, post stratification is used to try and reduce the non-response error. Strata are constructed according to a combination of sex, age, H region (local labour region) and civil status.

Statistics on children

The original sample has been redefined to apply to a population of children. Children included in the sample for statistics on children are those that are included in the families of the sample persons, on the basis of what is reported by the sample person. Observations can thus be taken from interviews with either the mother or the father.

Sampling and non-response errors

All sample surveys are marred by random errors, which are due to the random sample draw. The size of the error depends on which sample design and sample size have been used.

All selected persons in a random sample cannot or do not want to take part in the survey. Non-response reduces the size of the random

sample and therefore the random error is larger than it would otherwise be. If non-response is random, it does not necessarily influence the results. But non-response is often more or less selective. A low response rate within certain important groups can distort the survey results.

Non-response is generally lower among persons with children than among those without children. When recalculating the sample to a total level, therefore, an overestimation of the number of children is obtained systematically. At the same time, the higher response rate among families with children gives a comparatively good reliability regarding different breakdowns within the population of children and families with children.

Estimations

The three surveys LFS, LCS and HEK all have a more or less complicated structure.

LFS has stratified samples (independent random sample in strata) with after-stratification (and panels).

LCS has a systematic sample (approximate independent random sample) with after-stratification

HEK has a stratified sample of the network sample type.

The Parents Survey has a simpler design with stratified independent random sampling in the younger group and independent random sampling in the older group.

Proportion estimation

The following estimator for a proportion is applicable for a simple independent random sample of adult sample persons, where this sample is being used for the estimation of the proportion of children. For such an estimate, information must be procured on the number of children a sample person has.

$$p = \frac{\sum_{i=1}^{n} \frac{1}{\pi_i} a_i}{\sum_{i=1}^{n} \frac{1}{\pi_i} m_i} \qquad (1)$$

a_i = number of children in a family i with a specific property

m_i = number of children in a family i

n = random sample size (number of families with children)

π_i = inclusion probability for family i

N = number of families with children in the population frame

"Family i" is chosen through "sample person i" The inclusion probabilities in LCS and HEK are reduced because one family can be chosen in several ways (i.e. through both cohabiting parents). That one family could, in reality, be chosen twice is unlikely and is disregarded.

The estimator for the total number of children with a particular characteristic is made up of the numerator in (1). As the random denominator is not included, the variance calculation is simplified. Furthermore, it can be said that the inverse of the inclusion probabilities, the "adjustment coefficients" (disregarded from the reduction according to overleaf), are general for everyone in the survey, i.e. even those without children.

Estimations of variances for proportions and totals

When estimating a variance for a proportion, attention must be paid to the varying number of children per family. The variance can be expressed for a simple independent random sample as,

$$v(p) = \frac{1-f}{n} \frac{\sum_{i=1}^{n} \left(\frac{m_i}{\pi_i}\right)^2 \left(\frac{a_i}{m_i} - p\right)^2}{\left(\overline{\frac{m_i}{\pi_i}}\right)^2 (n-1)} \qquad (2)$$

where $f = \frac{n}{N}$

Note that the ratio in brackets refers to the mean in the denominator.

The variance in formula (2) can therefore be used with a simple independent random sample. But the different surveys have a more complicated design and the estimation of variances is in general much more complicated (at least for LFS and HEK).

The variances are used for the calculation of a confidence interval.

Confidence interval (95%):

$$p \pm 1{,}96\sqrt{v(p)}$$

Labour Force Surveys

Standard deviation for LFS has not been calculated. It can be estimated that the confidence intervals are roughly half those which apply for similar estimates from LCS.

HEK

The HEK is a network sample in which variance estimations for proportions can be made for all persons in the household and therefore even for children. Here, the ordinary HEK system for variance estimation has been used[1]. This gives a solution for the ratio estimation problem and also gives a well-defined estimator, which takes into account other "design" properties. A 95% confidence interval is given for a number of central estimates. The confidence interval is given in direct connection to tables in the chapter on the family economy.

Living Conditions Survey

Within the LCS system, a possibility for correction for the "network effect" has also been introduced and variances are calculated using a ratio estimate[2]. A 95% confidence interval is given for a number of central estimates in connection with tables in different chapters.

FU

A confidence interval is calculated for the FU with known formulae for stratified probability sampling.
The following formula is assumed.

$$P \pm d \cdot k \sqrt{\frac{P(100-P)}{P_{ng/n}}} \qquad \text{where}$$

P is the estimated percentage
K is the constant that gives the confidence level
D is the design effect that is calculated for every municipality
$P_{ng/n}$ is the reported group's percentage share of the whole sample in the municipality.

[1] Statistics Sweden, Extrakt system
[2] Statistics Sweden, Tanja system