EuReporting

Official statistics in Denmark: Socio-economic microdata for research

Bo Møller and Lone Solbjerghøj (Statistics Denmark),
Franz Kraus (University of Mannheim, MZES)

Official Social Surveys in Europe.
Edited by F. Kraus and G. Schmaus

EuReporting Working Paper No. 24

Towards a European System of Social Reporting and Welfare Measurement

A TSER-Project Financed by the European Commission

Subproject
Access to Comparative Microdata

Mannheim Centre for
European Social Research
(MZES),
Eurodata Research Archive

and

Centre for Populaton, Poverty
and Public Policy Studies
(CEPS),
Panel Comparability Project
Paco

Mannheim 2001

Editors' note:

'Official Social Surveys in Europe' is a series of commissioned working papers whose major aim is to provide basic information on the availability of and access to major official survey data in Western Europe relevant for problem-oriented social reporting on Europe.

The working papers provide

- background information on data infrastructures;
- a summary of survey profiles and access conditions for a set of major official social surveys;

The papers are supplemented by detailed, standardised survey documentation made available on the Internet at the home-page of the subproject at the Mannheim Centre for European Social Research (MZES)[1].

In addition to common documentation items, information also covers a set of major classification variables with respect to the availability and concepts/definitions. In compiling this information, authors could usually use as a point of reference meta-information made available to the project by Eurostat, Directorate E.

The information system on the Internet can be queried either to retrieve blocks of information for a single survey, or to compare single items (such as definition and measurement of household relationships) across a set of surveys.

---

[1] Currently accessible at http://www.mzes.uni-mannheim.de/projekte/mikrodaten/comi.htm

1. ORGANISATIONAL AND LEGAL ASPECTS OF THE SYSTEM OF OFFICIAL STATISTICS .. 1

2. THE STRUCTURE OF THE STATISTICAL SYSTEM 3

3. MAJOR SURVEYS IN THE FIELD OF SOCIAL STATISTICS 4

3.1 THE HOUSEHOLD BUDGET SURVEY.. 5

3.2 THE DANISH LABOUR FORCE SURVEY 10

4. COMPREHENSIVE SOURCES: INTEGRATED MICRODATABASES FOR RESEARCH 14

4.1 THE FERTILITY DATABASE 15

4.2 THE INTEGRATED DATABASE FOR LABOUR MARKET RESEARCH (IDA) 15

4.3 THE REGISTER OF SOCIAL RESEARCH (SFR) 15

5. COMPARABILITY ISSUES... 16

6. ACCESS TO MICRODATA FOR RESEARCH: MODES AND CONDITIONS............................... 17

7. LITERATURE AND ONLINE DOCUMENTS: 20

8. APPENDIX: SURVEY DESCRIPTIONS IN SUMMARIZED FORM 21

8.1 FORBRUGSUNDERSØGELSEN 1998 22

8.2 LABOUR FORCE SURVEY 2001 24

Official statistics in Denmark: Socio-economic microdata for research[2]

In Denmark's system of official statistics the collection of information via surveys has increasingly lost importance over the last three decades. Today, most of the required information - particularly in the field of social statistics - is collected from administrative sources. Survey response problems and scarcity of resources have initiated a gradual move from surveys to the use of administrative sources in many other countries, too. Although this series of working papers deals primarily with surveys, we decided to also include here information on the availability of microdata based mainly on administrative sources, though necessarily in a brief way. Extensive use has been made of material published by Statistics Denmark or articles written by staff members of the office[3].

The paper is divided into 6 substantive sections and an appendix: (1) organisational and legal aspects of the system of official statistics, (2) a short presentation of the data collecting system; (3) a presentation of the major socio-economic surveys conducted by Statistics Denmark; (4) an introduction to major microdata sources in the field of social statistics; (5) a discussion of comparability problems with the emphasis on the use of administrative sources; and (6) modes of access to microdata sources for research. In the appendix, standardised survey documentation is presented in a summarized form; more detailed topics can be quieried in comparative perspective via EuReporting's Internet pages[4].

1. Organisational and legal aspects of the system of official statistics[5]

In Denmark, official statistics is highly centralised. Statistics Denmark, the central statistical office, produces the bulk of official statistics; ministerial publications are limited to only few

---

[2] Texts and documentations for household budget survey and labour force survey by Bo Møller resp. Lone Solbjerghøj. Other sections by Franz Kraus, mainly based on material released by Statistics Denmark.

[3] Most prominently 'Statistics on Persons in Denmark. A register-based statistical system', ed. by Eurostat and Danmarks Statistik, Luxembourg: Office for Official Publications of the EC, 1995; 'Declaration of Contents', provided by Statistics Denmark on Internet at http://ww.dst.dk/dst/dst.asp?o_id=2460 (accessed July 2001).

[4] http://www.mzes.uni-mannheim.de/projekte/mikrodaten/comi.htm

[5] Major source: 'Presenting Statistics Denmark', http://www.dst.dk/data/893/893.pdf, accessed August 2001.

subject matters. The office is responsible for the coordination of all official statistics, supervises central public registers, and must be consulted by all public bodies with respect to design and content of newly established registers. Professionally, Statistics Denmark is an independent organisation. It is managed by a Director General (the National Statistician) and directed by an independent Board of Governors, consisting of the National Statistician and 6 persons with insight into social and economic conditions in Denmark. The Board decides on the work programme of Statistics Denmark, thus allowing the central statistical office to operate independently from direct government control. The Board is assisted by a number of subject-matter related advisory committees, in which both suppliers and users of statistics are represented. Administratively, Statistics Denmark is affiliated to the Minister of Economic Affairs, who is responsible for decisions on the budget, staff matters, and the appointment of the Board members[6].

The legal basis within which Statistics Denmark operates is formed by an act on data privacy ('The Act on Processing of Personal Data' of May 2000[7]) and the 'Act on Statistics Denmark'[8].

The *'Act on Processing of Personal Data'* substitutes the previous acts on public resp. private registers and defines rules with respect to processing data, the rights of data subjects, security, notification and supervision. According to this Act, Statistics Denmark is subordinate to supervision by the Data Protection Agency in all cases where personal data are processed solely for scientific or statitical purposes. In such cases, Statistics Denmark must first obtain an opinion from the Agency; before processing operations are started, the Agency must also be notified with respect to certain details of the operation. [9]:

The *'Act on Statistics Denmark'* establishes the office as the central authority of official statistics, settles the basics of governance, and defines the statistical office's major functions, the range of statistical topics to be collected as well as the obligation of data subjects to supply information for statistical purposes. The act stresses the need to reduce response burdens by stating that, as far as possible, use should be made of information already collected by other public bodies and allows Statistics Denmark access to data from all public administrative registers in Denmark. The 'Act on Statistics Denmark' has been designed as a framework, giving considerable leeway regarding the development of enumeration programmes, the specification of contents and the data collection methodology. Major players are the National Statistician, the Board of Governors, the advisory committees and the

---

[6] The chairman of the Board, i.e. the National Statistician, is appointed by the King upon recommendation of the Minister of Economic Affairs.

[7] Act No 429 of 31 May 2000. This Act implements the 1995 Directive of the EC (No 95/46/EC) on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[8] Act No. 1182 of December 21, 1992, as subsequently amended by section 1 in Act No. 295 of May 2, 2000. Major parts of this act go back to the Act on Statistics Denmark of 1966.

[9] Act on Processing of Personal Data (Act No. 429 of 31 May 2000), sections 43 to 45.

Ministry of Economic Affairs. Competencies are clearly defined in terms of ultimate responsibilities, but interest mediation through consultation has been institutionalised in various forms. Legal provisions with respect to procedures are limited to general principles, leaving their implementation up to the organisations involved. In contrast to systems based on strict legislation at the level of single enumerations, as it is the case with the European Union and Germany, the Danish system thus shows great flexibility with respect to adapting to new data needs and innovations in data capture.

2. The structure of the statistical system

In the 60ies, Statistics Denmark decided to make use as intensively as possible of administrative sources in order to reduce response burdens. Since then, backed by statistics legislation, the statistical system has gradually been transformed, and today administrative registers are the office's principal data source. Only in the case of sectoral business statistics is a wide range of data collected from enterprises with the help of questionnaires. General business statistics, covering all sectors, and social statistics are today predominantly based on administrative register data. The Census of Population, for example, was carried out entirely on the basis of administrative sources for the first time in 1976, and the Census of Housing in 1979. Since then, the Census of Population and Housing has been carried out this way every year.

The use of administrative data for statistical purposes has profoundly been facilitated by the introduction of unique identifiers (identification numbers for persons, for real esteates, and for businesses). The uniqueness and unambiguity of these identifiers allow Statistics Denmark to directly link information across statistical units (such as persons with workplaces) and/or subject matters.

Register sources available to Statistics Denmark comprise both basic registers on units (such as the Central Population Register, the Central Register of Buildings and Dwellings, the Central Business Register) as well as a huge variety of specialised registers resulting from administrative processes (such as the Central Register of Salary Information, the Central Register of Final Assessments of the Central Customs and Tax Administration, Register of the Directorate for Employment and Vocational Training regarding Labour Market Measures, etc.). These sources are complemented by so-called 'classification modules' with general rules (typically conversion keys) for the grouping of units according to statistical concepts. Examples are the Household and Family module (used for grouping persons into families and households according to definitional requirements), the Educational Classification Module (used for grouping people according to educational characteristics), the Employment Classification Module (used for grouping people according to their usual employment characterisitics).

Source data and classification modules are combined to produce statistics in accordance with the work programme and the statistical concepts/definitions to be applied. Resulting data are stored in a new file - a so-called 'statistical register'. Traditionally, these files are organised according to subject matters (covering only those variables immediately needed for the respective topic), and in the case of data linkage across sources identification keys are deleted after the successful creation of the output data. More recently, as a result of increasing needs for data composition across different registers, comprehensive statistical registers were created (cutting across traditional subject matter divisions) with identification keys no longer deleted (the so-called 'integrated registers'). These integrated registers are obviously more efficient than the traditonal ones with respect to analytical use, data integrity and up-dating, but involve additional measures to meet the high security standards of the security regulations for Statistics Denmark. In principle, data can be linked not only across sources ('horizontal integration'), but also over time ('vertical integration'). Vertical integration, i.e. the creation of longitudinal data through linkage is governed by legal provisions and the actual content of the working programme as decided by the Board of Governors. In practice, most of the statistical registers refer to a certain point in time, while only a few of them also provide flow data.

Statistics Denmark provides detailed and standardised documentation of sources and responsibilities (administrative information), contents, time (periodicity, data coverage, up-dating), accuracy, and comparability, both for register data and survey data online at http://ww.dst.dk/internet/startuk.htm[10].

As mentioned above, in social statistics and general business statistics administrative sources have replaced the conduction of surveys with a few exceptions only. In both fields the use of surveys is primarily restricted to two purposes: supplementing administrative sources or meeting the demands of Eurostat in those cases where the use of survey instruments is part of the EU statistics legislation.

3. Major surveys in the field of social statistics

Apart from a few regular surveys for internal and external customers of Statistics Denmark, such as the monthly *Omnibus Survey*, the major surveys in the field of social statistics are the *Danish Labour Force Survey* and the *Household Budget Survey*. Both are related to data requirements of the European Union. While the Household Budget Survey is coordinated with the requirements of Eurostat at a voluntary and an ex-post basis, the Danish Labour Force

[10] Select 'Guide to Statistics'

Survey falls under EU statistical legislation. Two other European collection programmes, the labour cost survey and the earnings structure survey, are mainly based on administrative sources, supplemented by questionnaire-based information where required. The *Census of Population and Housing*, for which there is a European committment with respect to tabular data, was completed in Denmark in 1970 for the last time on the basis of questonnaires. In 1981 Denmark, as the first country of the world, conducted the entire Census solely on the basis of register data. Since then the Census has been abolished and data for most of the census topics are collected from registers on an annual basis. The *European Community Household Panel* is conducted via sample surveys and according to the European standard questionnaire. Its conduction, however, has not been made a task of Statistics Denmark; instead, the Danish Institute of Social Research is collecting these data on behalf of the European Commission.

The following section is therefore limited to the two major socio-economic surveys of Statistics Denmark: the Danish Labour Force Survey and the Household Budget Survey.

3.1 The Household Budget Survey[11]

3.1.1 History

The household budget surveys (HBS) have a very long history in Denmark. The first survey was conducted more than 100 years ago. The earlier surveys did not cover all household types, but just special groups – that is in the later years just employees, e.g.. From 1976 all household types were covered. Up to 1994 the survey was conducted at intervals of 5-6 years: 1976, 1991 and 1997 –, but since 1994 the methodology of the survey has been changed, and the survey is now conducted continuously with annual updates of the results.

3.1.2 The sample

The HBS is a rather expensive survey. For budgetary reasons the sample size has been limited to 1.500 addresses each year. This number has been found to be too small to form the basis of more detailed results. Therefore the data collected in 3 subsequent years are accumulated to form a gross sample of 4500 addresses.

The participation in the survey is voluntary for the households. And, at the same time, the survey participation demands a lot from the households. Thus the survey – as does other countries' HBS - suffers from a rather high non-response rate. In Denmark the non-response rate is about 1/3, even though Statistics Denmark offers the households several kinds of 'compensation':

- A gift chosen from a list comprising a coffee maker, a camera, a radio, etc.;

[11] Text and standard documentation (cf, appendix and EuReporting database in the Internet) provided *by Bo Møller, Statistics Denmark*

- The participation in a lottery. Each month a prize of DKK 10.000 is drawn among the about 1000 households who participate in one month;
- To all households who have used more than 1 diary a bag of candies is given – the main idea is to make the whole household, including older children interested in diary-keeping and the survey as such.

The basis of the sample are the administrative registers of all inhabited addresses in Denmark. The visit of the interviewers to the sampled addresses establishes the actual sample of households as an economic entity. The sample is drawn in geographical clusters grouped around the interviewers working for the survey. It is ensured that all big cities are covered and that Denmark is covered in a reasonable way according to rural/urban environment and main parts of the country. The households in the total sample are distributed evenly round the year so that seasonal variations,e.g., are neutralized.

The fieldwork as such is contracted to an external partner (at present the Danish Institute for Social Research), as Statistics Denmark does not have its own corps of interviewers for interviewing in the field. The total responsibility for training the interviewers and all materials used remains with Statistics Denmark. The sampled households are visited two times. During the first visit the interviewer establishes the precise composition of the household in cooperation with the household. The interviewer gives instructions on diary keeping and conducts the first part of the interview (as computer-assisted personal interviewing – CAPI –, using Blaise software). The following two weeks the households record all expenses in the diary. During the second visit of the interviewer the diaries are collected and the second part of the interview is conducted.

3.1.3 The data sources

Three very different data sources form the basis of the HBS:

- The interview with the households;
- The diary kept by the households;
- The information found in the administrative registers.

The interview is very long. The main topics are:

- Background information on the household and its members;
- Fixed expenses;
- Major expenses on durables, etc.;
- Use of educational, health and child care services;
- Pension schemes
- Certain incomes.

The reference period of the interview is the 12 months prior to the day of the interview.

The diaries are open – that is the households record in their own words the sort of expense and the amount. Additional information is included on the kind of food preservation used, on clothes and shoes according to sex and age as well as information on whether the expenses were made in Denmark or abroad – this latter information is used when the share of indirect taxes in consumption is imputed. The diary covers a period of 2 weeks starting the day after the first interview. Each household receives one 'big' diary and can use several small pocket-size diaries as well.

In cases where the diary includes information on items covered by the interview the interview-based information is used – that is, the diary-based information is deleted by the bureau after it has been controlled.

The information collected directly in the households is afterwards supplemented with information based on administrative registers. The linking is direct as Statistics Denmark has information on the unique personal identification number of each person in the survey.

Several types of register-based information are used:

- The register of incomes is the most important one. It is based on the tax records of each person and comprises very detailed information on different income types, on taxes, on interest paid, etc.
- Register-based information on the socio-economic status of each person.
- Information on demography
- Information on the educational level
- Information on the dwelling (size, age, etc.)

The register-based information is used in three different ways:

- Directly as variables in the survey
- As background variables on the household
- As the basis of the calculation of weights.

The use of register-based information has two main advantages:

- It saves time as the office does not have to include those topics in the interview where access to the information via the registers is possible
- It can give better information as households often do not remember exactly which incomes they had or may not want to give precise information.
- On the other hand, of course, the use of register-based information has some limitations.
- The concepts used in the registers are determined by the administrative use. Therefore the statistical concept that should be applied not always corresponds to what the office wants.
- Black incomes, etc., are of course not available in the registers. Therefore the interview has a small set of income questions on income types (including black incomes), which cannot be found in the registers.

3.1.4 *The checking of data:*

The interview-based data are checked during the visit to the households. For each variable in the interview a minimum and a maximum are fixed. An amount outside this range gives a warning to the interviewer who then checks the information once again with the household. In addition, other checks in the interview programme ensure that no illegal characters or illogical answers can be typed in.

The *diaries* are coded and checked in the bureau. This is done by using Blaise software. As with the interviews, all amounts below or above some fixed limits give a warning to the coders. The *register-based information* is not checked as such, but it is ensured that the information exists for all persons. This can be a problem in special cases – for persons who died after the date of the interview, who have emigrated, etc.

3.1.5 *Data treatment*

The data from the different sources are brought into the same edp format by using a very detailed coding system. For the consumption part the codes build on the COICIOP nomenclature. This process also takes care of the problem of the different reference periods of the data sources: one year back for the interviews, two weeks for the diaries and a calendar year for the register-based information.

Rents for owner-occupied dwellings are imputed as a part of consumption and as income as well. The basis for the imputation is information on the rent for similar rented dwellings found in the general statistics on rents.

The amount of indirect taxes in consumption is imputed by using the detailed consumption figures and the tax rules, etc.

For educational, health and child care services the indirect transfer from the public to the households is imputed as the difference between the direct payment of the households and the total cost of the service. This information on indirect transfers is used for the analysis of the economic relations between the households and the public in total, that is, direct and indirect transfers to and from the households.

For each household the identity: incomes = uses of incomes is established. The difference between the two sides is interpreted as positive or negative savings of the households.

For a smaller number of households this relation may not seem plausible – that is, the households can, e.g., have a much higher consumption than can be 'justified' by an analysis of the incomes and other variables. For these households corrections may be introduced or the households may be removed from the survey.

As mentioned above, the survey data are collected over a period of 3 years. The values for each household are adjusted to cover the situation in the year in the middle by using the quarterly national account figures. By using these figures for the adjustment, the changes in price levels as well as in quantities are taken care of simultaneously.

3.1.6 Weighting

Before publication, all survey data have to be weighted to adjust for different non-response rates, e.g. For this purpose a rather sophisticated poststratification programme is used.

The background for the poststratification is the register-based total for Denmark as a whole. The factors in the stratification are: household size and composition, household income, socio-economic status of the head of the household, degree of urbanisation, and ownership of a dwelling.

3.1.7 Publications:

Each year a publication on the main figures from the survey is produced. This publication is then supplemented with a special publication on a chosen subject. In the Statistical Yearbook, main results are published.

3.1.8 Other types of access to data:

For Eurostat a special file is constructed. Although the coding system used in Denmark is COICOP-based, a few differences between the national and the Eurostat codification exist, and Eurostat uses some background variables that are not used nationally. In the Internet (http://www.statistikbanken.dk), detailed tables can be found. Statistics Denmark can – against payment - produce special tabulations and provicre other services. for all kinds of users. Researchers can have access to the micro files under some special conditions and against payment.

3.1.9 Documentation:

In the yearly publication, the method is described. A more detailed description can be found in the book: 'Forbrugsundersøgelsen – Metodebeskrivelse - Fra dataindsamling til offentliggørelse, Statistics Denmark 1999'. In the section 'Declaration of contents', Statistics Denmark also provides a short-cut documentation in English (visit http://www.dst.dk, respectively http://www2.dst.dk/internet/varedeklaration/en/V01079.htm)

3.2 The Danish Labour Force Survey[12]

3.2.1 Purpose and History

The primary purpose of the Labour Force Survey is to give a description of the labour market attachment of the Danish population. In this description the population is grouped into the categories employment, unemployment or inactiveness (not in the labour force, e.g. students or early retirement pensioners).

Labour Force Surveys are carried out in many countries, and common concepts and guidelines are used, which make the Labour Force Surveys suitable for international comparisons.

In the period 1984-1993 the Danish Labour Force survey was conducted as an annual survey. Since 1994 the Danish Labour Force survey has been carried out as a continuous survey.

3.2.2 Scope and Content

Main variables:

- Labour market attachment

In the Labour Force Survey (LFS) the population is grouped into two main categories: persons in the labour force and persons not in the labour force. Furthermore, persons in the labour force are categorized as either employed or unemployed. Conscripts are considered as employed. The grouping of respondents into the above-mentioned groups is based on their labour market attachment in the reference week, which is the survey week the answers refer to. The definitions used in the Danish LFS are recommended by the International Labour Office, ILO, and Eurostat has interpreted and operationalized the guidelines that apply to the LFS in the EU member states. *Employed* are all persons who worked at least one hour against payment in the reference week or who are self-employed or family workers. Persons who are temporarily absent should have an assurance of being able to return within 6 months. *Unemployed* are all persons who are not employed, have actively been seeking work in the 4 weeks before the interview, and are available for work within 2 weeks.

- Main activity, activity a year ago, education

Furthermore, data are collected on the main (usual) activity – both for the situation in the reference week and for the situation one year prior to the interview. Data are collected on completed education as well as on the education activities the respondents have been engaged in within the last 4 weeks.

---

[12] Text and standard documentation (cf, appendix and EuReporting database in the Internet) provided by *Lone Solbjerghøj, Statistics Denmark*

- Characteristics of employment

For the employed variables are collected that describe their main jobs, and, if the person holds a second job, some selected variables for this job are collected, too:

  - Professional status
  - Economic activity
  - Distinction between full-time/part-time
  - Duration of employment
  - Number of hours usually worked
  - Number of hours actually worked
  - Work-hours pattern
  - Search activities for another/extra job
  - Characteristics of the non-employed

The non-employed are asked about their availability for employment and search activities as well as for a few aspects of the job they want.

Ad-hoc surveys

Since 1999 data on specific topics have been collected on an ad-hoc basis. The collecting of data is legalized by Council Regulations. Eurostat publishes the results. The following surveys were conducted:

  - 1999: Accidents at Work and Occupational Diseases
  - 2000: Transition from School to working life
  - 2001: Length and patterns of working time

3.2.3 Data sources and the use of administrative and statistical registers

Survey design

The Danish Labour Force Survey has been carried out in its present form with a continuous reference week since 1994. The quarterly sample size is 16,665 persons between 15-74 years. The sample is divided into 13 samples of equal size. The persons in each sample are interviewed with reference to a specific week (the reference week). The Danish Labour Force Survey is a rotating panel survey. Approximately one third of the panel is replaced each quarter, one third is interviewed again the following quarter and yet another time one year later. This makes it possible to compare the answers given by the same person in two consecutive quarters and again one year later. E.g., one third of the sample interviewed in the first quarter 2001 is interviewed for the first time in this quarter and will be interviewed again in the second quarter 2001 if possible, and yet a third time in the second quarter 2002 if possible.

Interview data

Since 1986 the survey has mainly been conducted via telephone interviewing. Persons who could not be reached by telephone received a mailed questionnaire instead. According to the regulation, only the classification of the labour status has to be based on interview data, for other variables register data may be used if available.

Register-based data

Some of the variables concerning the labour force are deduced from register data. Furthermore, all data for population groups 0-14 and 75+ are added from registers.

- Varables taken from registers:

Demography: The information on the demographic background of the respondent is taken from the population register and includes:

  - The position in the household
  - Sex
  - Age
  - Marital status
  - Nationality
  - Years of residence
  - Country of birth

Education: Collecting data on attained (vocational) education and the coding of the information is very burdensome; therefore these data are derived from registers whenever possible/therefore, registers are used whenever possible, and interview data are only used as a supplement. If a respondent's education was completed recently, it might not be in the register yet, therefore in those cases where the respondents' education is registered they are asked if they have completed their education recently (approx. within the last 2 years). Education completed abroad is not included in the register, thus Statistics Denmark asks probe questions to check if the respondent has completed some kind of relevant education in another country.

Other variables of the basic module: Data in the survey are register-based in some other cases as well, e.g., registration at a public employment office .

Variables of the ad-hoc-module: In some cases it is possible to include register variables for the ad-hoc surveys, too. For the 2000 survey on "Transition from school to work" it was even possible to base the survey exclusively on register data - mainly on the Danish educational register system.

- Extension of survey population: the children and the elderly

Currently register information for a sample of persons aged 0-14 years and 75 years and more is added to the survey population in the LFS files that are sent to Eurostat. A fixed sample size of 3,159 persons aged 0-14 years and a fixed sample size of 1,225 persons aged 75 years and more is drawn fom the population register.

3.2.4 Use of administrative registers in the survey design

The sampling process:

In the sampling selection phase persons that were registered as unemployed two quarters prior to the survey quarter are selected with a higher probability than their relative proportion of the total population is. This is done to ensure enough observations among the unemployed. Thus approximately one third of the sample is drawn from the Unemployment Register, whereas two thirds of the sample are drawn from the Population Register, as this group was not registered as unemployed two quarters prior to the survey quarter. The sample units are individuals.

The weighting process:

- Two weighting procedures:

As mentioned above, in the sample selection phase persons that were registered as unemployed two quarters prior to the survey quarter are selected with a higher probability than their relative proportion of the total population is.. Due to the facts mentioned above and due to a justified assumption that there are different non-response patterns, persons previously registered as unemployed and persons not registered as unemployed two quarters prior to the survey quarter are weighted in different strata. A total of 53 strata are used in the poststratification phase to take the non-response into account. 18 of these strata are related to persons previously registered as unemployed, whereas the remaining 35 strata are related to persons not registered as unemployed two quarters prior to the survey quarter.

- Persons previously registered as unemployed

For people that are selected from the Unemployment Register the current register variables used for poststratification are:

  - age (3 groups),
  - education (3 groups), and
  - gross income (2 groups).

- Persons previously not registered as unemployed:

On the other hand, for people who are selected from the Population Register the current register variables used for poststratification are:

  - age (3 groups),
  - gross income(3 groups), and
  - industry (4 groups).

3.2.5 *Dissemination and application of the statistics*

Each quarter a *News Release* and the *Statistical News* are published, containing the main results from the survey. In the *Statistical Yearbook* and the *Main Indicators* main results are published with notes in English. Eurostat publishes the results for all member states in the publication *Labour Force Survey* and receives a data file including persons under 15 years and over 74 years. Tables are available in the database New Cronos. In the database, *http://www.statistikbanken.dk*, detailed tables can be found. It is planned to make a version in English. Statistics Denmark can – against payment - produce special tabulations and provide other forms of services for all kinds of users.

Researchers can have access to microdata under some special conditions and against payment.

4. Comprehensive sources: integrated microdatabases for research[13]

Statistics Denmark maintains a variety of register-based microdatabases relevant for social research. Most of these registers are, however, limited to particualar subject areas. In addition to these sources, Statistics Denmark has also established a number of wide-ranging cross-subject databases, generated for research purposes. For this aim, statistical registers from a variety of different fields of statistics are directly linked. In contrast to subject-area related statistical registers, identifiers are not deleted afterwards.[14]. Thus, additional linkages can be performed by Statistics Denmark in an efficient and flexibele way, making these databases highly valuable for a variety of research purposes.

The Fertility Database, the Integrated Database for Labour Market Research (IDA), and the Social Research Register (SFR) are major examples. All of these databases allow for longitudinal analyses. Further examples, not covered here, are the Occupational Mortality Database, the RIMO model for simulations as well as The Children Register and the Register for Preventive Medicine[15]. Interested users are highly recommended to visit Statistics Denmark's Internet documentation 'Declaration of Contents' to get further information.

---

[13] Based primarily on Eurostat and Danmarks Statistik: 1995 as well as on the documents provided on Internet by Statistics Denmark (cf. 'Declaration of Contents' at http://www.dst.dk/dst/dst.asp?o_id=2460 (accessed July 2001).

[14] Cf. Statistics Denmark: 1999.

[15] Cf. Statistics Denmark's Internet documentation 'Declaration of Contents' and Statistics Denmark: 1999)

4.1 The Fertility Database[16]

The Fertility Database has been developed to support a socio-demographic analysis of fertility trends. It covers females aged 10-49, males aged 13-64, and all children with parental references. The database contains a large number of basic variables (demographic background, education, socio-economic condition, information on dwellings, information on family contexts) and a range of derived variables (age at first birth, parity, birth intervals, etc.). Time coverage differs somewhat across subject areas, ranging from 1980 (demographic information) to 1984 (educational information).

4.2 The Integrated Database for Labour Market Research (IDA)[17]

The main purpose of the IDA is to support research on a wide variety of labour market issues. It contains information on the population and its attachment to the labour force and links data on employees to data on employers at the level of establishments and firms. The database is longitudinal, allowing the monitoring particularly of year-to-year changes since 1980. More than 200 variables (basic and derived) are included (as of July 2001), covering persons (demographic background, education, employment and work experience, unemployment, income), jobs (occupation, work-scheme, earnings, seniority, recruitments/resignations), establishments and firms (business demography, employees and earnings, industry). The database has been estblished as a joint effort of the Danish Social Science Research Council and Statistics Denmark and is updated annually.

4.3 The Register of Social Research (SFR)[18]

The SFR has been developed to support research on the development of living conditions. It is a longitudinal register, covering a period of 10 years back in time (rolling period). It combines data selected from nine different sources, including a small selection of variables from the IDA. At present, the database includes more than 400 topics (basic and derived variables) with a clear focus on income. The database has been established as a joint effort of the Danish National Institute of Social Research and Statistics Denmark; it is up-dated annually. Due to changes in classification systems, comparability over time is somewhat hampered with respect to industry, occupation and socio-economic status.

---

[16] Cf. : http://www2.dst.dk/internet/k13/fertil98/fertil_1.htm and http://www.unece.org/stats/documents/1999/03/confidentiality/18.e.pdf

[17] Cf. http://www2.dst.dk/internet/varedeklaration/en/V01013.htm and http://www.unece.org/stats/documents/1999/03/confidentiality/18.e.pdf . Cf. also Statistics Denmark: 1996 for a description of longitudianl data and their analytical possibilities with respect to labour market changes.

[18] Cf. http://www2.dst.dk/internet/varedeklaration/en/V01016.htm and http://www.unece.org/stats/documents/1999/03/confidentiality/18.e.pdf

5. Comparability issues[19]

In Statistics Denmark's 'Declarations of Content' three aspects of comparability are discussed: international comparability, comparability over time, and comparability with other sources.

Concerning the two *surveys* described in section 3, both have increasingly been subject to considerable international harmonisation efforts by the European Union.

This holds particularly for the *Danish Labour Force Survey*[20] which is based on detailed EU statistics regulations, the most recent version[21] of which has been gradually implemented in the Member States. Denmark expects that this regulation will be fully implemented by 2003. Although in Denmark certain background variables are included from registers and despite all the differences in survey details[22] between the Member States, the Danish Labour Force Survey is based on a common methodology and on common definitons, thus allowing to derive cross-nationally comparable results since 1994. Concerning comparability over time, the survey has been conducted according to similar guidelines since 1984. However, minor changes in the definition of unemployment are reported for 1992, 1994, 2000 and 2000, a change from a yearly survey conducted in spring to a continuous survey in 1994, and a major questionnaire change for 2000. Comparability across sources within Denmark is reported to be more limited, although consistency across sources is reported with respect to broad trends.

The *Household Budget Survey*[23] is not based on EU legislation; comparability across the Member States is instead based on the harmonisation of output and therefore more limited as in the case of the Labour Force Survey. Again, some of the variables are derived from administrative sources, making the survey very rich in information detail compared with household budget surveys of other countries.[24] Concerning comparability over time, statistics have been highly comparable since 1994, while widely changing definitions and classifications limit comparability to earlier periods. Comparability across sources is reported to be most limited with respect to population statistics, due to different definitions of the household.

---

[19] This section draws mainly on Statistics Denmark: 1995 and on Statistics Denmark's Internet documentation 'Declaraton of Contents'.
[20] For comparability issues cf. Statistics Denmark: 'Declarations of Content', at http://www.dst.dk/dst/dst.asp?o_id=2460, Labour Force Survey.
[21] Council Regulation (EC) no 577/98 is the most recent regulation
[22] One major difference relates to unemployment, where in Denmark contacts with a public employment office, local authority or unemployment fund within the past 3 months are considered as active job search, while according to the guidelines of Eurostat the framework should be limited to 4 weeks (cf. Statistics Denmark: Declaration of Contents - Labour Force Survey). This introduces a bias towards higher participation rates as well as higher unemployment rates in Denmark.
[23] For comparability issues cf. Statistics Denmark: 'Declarations of Content', at http://www.dst.dk/dst/dst.asp?o_id=2460, Household Budget Survey. Cf. also Eurostat:
[24] Cf. Eurostat: 1993.

Concerning register-based microdata, comparability issues are even more difficult to assess at a general level. In international discussions, the major arguments against the use of administrative registers are concern their validity, stability and quality. Administrative sources, in the first line, are geared to administrative requirements. Register concepts and statistics concepts do therefore not necessarily coincide; changes in legislaton and administrative processes may hamper the stability of registers over time, and codes needed for linking data may be incomplete or subject to administrative changes, as, for example, in the case of renumbering the houses in a street. Statistics Denmark has continuously been working at gradually reducing the adverse impacts of these factors on the quality of statistical registers. Statistics Denmark has developed a variety of strategies to cope with these problems: negotiations with administrations to take account of statistical needs and to improve the quality of registers, the establishment of general classification modules to ensure the consistency of grouping operations, the reduction of inconsistencies through defining central variables[25], and the establishment of accounting systems in addition to national accounts[26], the integration of registers across subject-matters and/or over time, the linkage of microdata from register data and survey data for common topics, the establishment of a global metadatabase, etc.

Maintaining and improving the qualitiy of registers is a continuous task, and the more intensively registers are used for statistical purposes, the more can they be improved.. Similarly, the use of data for scientific purposes adds to the gradual improvement of data quality. The establishment and maintenance of comprehensive microdatabases for scientific research must be viewed from this perspective, too.

6. Access to microdata for research: modes and conditions[27]

Statistics Denmark has developed a number of standard products to disseminate statistical information to a wide variety of users. In addition to the dissemination of print and electronic publications, various services are offered at the level of tabular data to meet more specific data needs. Statistics Denmark offers free access to its statistics databases[28] with a large number of detailed multi-dimensional tables which can be tailored by the users to meet their needs. In addition, special services can be commissioned, covering model calculations as well as the linking or re-grouping of data to provide users with new data not contained in the office's standard products.

---

[25] Cf. Statistics Denmark: 1999
[26] Cf. Bulmann et. al.: 2000 for the introduction of Labour Accounts.
[27] Cf. Statistics Denmark: 1995 and information on Internet (http://www.dst.dk)
[28] http://www.statistikbanken.dk/

The provision of data for research is an explicitely stated obligation of Statistics Denmark. Very often, however, the provision of tabular data is insufficient, and researchers need access to microdata. 'Current legislation permits disclosure, to a wide extent, of personal data for scientific purposes.' (Anderson: 2001, p.2). As the office has a huge potential of information on individual units at its disposal, special guidelines have been developed to ensure that the very high standards of data security and confidentiality are not violated. For this purpose, Statistics Denmark has developed a variety of access modes, all of them devised to keep dissemination of data outside of Statistics Denmark at the very minimum.

In addition to commissioning services to Statistics Denmark for model calculations, the following modes of access to microdata are offered for research purposes:

- Direct remote computing:[29]

  Since March 2001, Statistics Denmark offers researchers the possibility to make direct batch runs on approved microdatasets from their external workplaces.[30] Statistics Denmark provide research data at dedicated servers within Statistics Denmark and users communicate with these servers via Internet. Communication is encrypted and access controled through passwords and individual hard keys (tokens). Printing and data transfer options are not available. Printouts are sent to the researchers by e-mail. Operations are stored in a log-file and checked randomly by Statistics Denmark.

  The reseach environment must apply for an authorisation which is always limited to specified research projects. A project description must be submitted, stating the research goals and the data needs. Usually, only samples of data are provided. If a researcher needs access to whole populations, variables must be selected. Particularly sensitive data (such as crime data) are not included in this scheme, and data on enterprises are assessed very carefully to avoid any problems of unlawful disclosure. Applications are assessed by a steering group consisting of the Directors of Statistics Denmark. The scheme has been approved by the Board of Governors and will regualarly be assessed by them. Researchers which are not granted access authorisation under this scheme can apply for other modes of access.

- Research at the premises of Statistics Denmark (accomidation scheme):

  Researchers are sworn-in as regular staff members are, and work has to be carried out in-house. This service includes the linking of academia-based sample survey data to official microdata through Statistics Denmark;

- Provison of model and study datasets for external use:

  In exceptional cases, Statistics Denmark might supply very small samples of microdata (eg., 1000 records) for research outside the office's premises.Usually data is suppled for

---

[29] We would like to thank Otto Andersen (head of Statistics Denmark's division 'Methods and Research') for his support. The following part is based on a recent (not yet published) article of Otto Andersen on recent changes regarding access to microdata of Statistics Denmark. Cf. Andersen: 2001.

[30] Prior to March 1, 2001 users hat to submit their programme setups to Statistics Denmark in electronic form; after scrutinizing the setups for forbidden applications, the job was run and the results were returned to the user via e-mail

the purposes of developing and testing computer programs at reseachers' work places. According to Andersen (2001), only a few model datasets have been created so far.

In addition to model datasets, Statistics Denmark has also prepared some study datasets for use in teaching. These study datasets are longitudinal, following a few thousand persons over time, and restricted to a subset of variables. Data are scrambled/grouped to avoid unlowful disclosure, but basic characterisitcs of the data have been preserved.

In any case, microdata are provided only in an anonymised form, and only those variables are provided which are deemed essential for meeting the research purpose. In all three cases, the researcher has to submit a research plan, specify the data needs and sign a contract with Statistics Denmark after approval through a special committee. Each case is covered by a different contract. In all three cases, however, the contract limits the period of use, declares that all data should be regarded as confidential data according to the Penal Code, and forbids the copying of data[31].

Users must compensate Statistitics Denmark for all resources used. In order to reduce start-up costs of new projects, the office – with the approval of the Data Protection Agency and financial support through external funds - has established a number of semi-fabricated microdatabases for research. Today, Statistics Denmark is the principal supplier of data for many research projects.

[31] The obligation to submit written texts to Statistics Denmark for scrutinizing prior to publication has in the meanwhile been abolished.

7. Literature and online documents:

Andersen, Otto (2001). Access to micro data from Statistics Denmark.. Unpublished manuscript. (Contact: Statistics Denmark, division Research and Methods)

Buhmann, Brigitte and W. Leunis, L. Solling, A. Vuille, K. Wismer (2000). 'Labour Accounts: A Step Forward to a Coherent and Timely Description of the Labour Market.' Paper presented to the 2000 meeting of the Siena Group on Social Statistics, Maastricht, 22-24 May 2000.

Danmarks Statistik (1999), 'Forbrugsundersøgelsen. Metodebeskrivelse. Fra dataindsamling til offentliggørelse'.

_____ (1998). 'The Fertility Database. Guide to extracting data from the Fertility Database'. Accessed at http://www2.dst.dk/internet/k13/fertil98/lbk_vejl.pdf

_____ (1995). 'Statistics on Persons in Denmark. A register-based statistical system'. Edited by the European Commission. Luxembourg: Office for Official Publications of the European Communities.

_____. 'Presenting Statistics Denmark.'
Accessed last August 2001 at http://www.dst.dk/data/893/893.pdf

_____. 'Declarations of Content.' Accessed last in August 2001 at http://www.dst.dk/dst/666

Datatilsynet. 'Act on Processing of Personal Data.'. Accessed last in July 2001 at http://www.datatilsynet.dk

Eurostat (1993). 'Family Budet Surveys in the EC: Methodology and Recommendations for Harmonisation. Luxembourg: Office for Official Publications of the European Communities.

Statistics Denmark (1999). 'Access to Confidential Data in an Integrated Statistical System'. Joint ECE/Eurostat Work Session on Statistical Data Confidentiality. Thessaloniki, Greece, 8-10 March 1999, Working Paper No. 18. Accessed last in August 2001 at http://www.unece.org/stats/archive/docs.e.htm

_____ (1999). 'Formation of central variables in a decentralised statistical system'. Joint ECE/Eurostat Work Session on Registers and Administrative Recordes for Social and Demographic Statistics. Geneva, 1-3 March 1999. Working Paper No. 25, prepared by Finn Spieker, Statistics Denmark. Accessed last in August 2001 at http://www.unece.org/stats/archive/docs.e.htm

_____ (1996). 'Changes in the labour market: Longitudinal data and analyses of persons and establishments.' UN Economic and Social Council. Statistical Commission and ECE. Conference of European Statisticians. 44th plenary session. Paris, 11-13 June 1996. Accessed last in August 2001 at http://www.unece.org/stats/archive/docs.e.htm

8. Appendix: Survey descriptions in summarized form

For details and comparative queries visit
http://www.mzes.uni-mannheim.de/projekte/mikrodaten/comi.htm

8.1 Forbrugsundersøgelsen 1998

Title
Forbrugsundersøgelsen

Parallel title
Household Budget Survey

Investigator
Danmarks Statistik

Subject items
The main content of the survey is detailed information on: incomes, consumption, the use of health-, education- and child care-services, the disposal over consumer durables, the coverage of pension schemes, etc.

Statistics domains
Demography, consumption, use of health-, education- and child care-systems, pension schemes, incomes, demographic characteristics, socio-economic status, education, housing conditions, durables in the households, etc.

Abstract (Study Purpose)
The HBS is a multi purpose survey.
Main uses are:
Internally: as one of the data sources for the national accounts and for the establishment of the weights in the price indices.
Externally it is used as an important source of information and research on living conditions, etc. and for panning purposes.
To this comes marketing purposes, etc.

Reference period
1999

Institutional universe
All private households (excluding institutions, etc.)

Geographic universe
Resident population

Type of source
Household survey (with some information on person level)

Study frequency
Permanently

Sample Size
4.500 addresses (1.500 each of the 3 subsequent years)

Sampling method
Approximately 30 geographical clusters are selected 'by hand' in a way that ensures:
That the greatest cities are covered,
That there is a reasonable distribution between the main parts of the country,
That there is a reasonable distribution according to the degree of urbanisation
That the clusters are made in a way that gives a reasonable way of transportation for the chosen good interviewers.

Total response rate
62,9 percent. – but in the subsequent data treatment process some further households are let out so the final figures builds on 61,8 percent of the sampled.

Class of survey
Household Budget Survey

Core Variables
gender, age, marital status, education, continued education, labour status, status in employment, working hours, occupation, industry, socio-economic classification, income, tenure status, housing amenities

Standard Codes
ISCO-88, ISCED, NACE R.1

First year of conduction
1994 in this form, but surveys with almost similar content have been conducted since 1976.

Conduction history
1994 -: annually

Data disseminators
Statistics Denmark, Forbrugsundersøgelsen

Statistical disclosure control methods
Removal of direct identifiers

Type of access
For scientific uses and for Eurostat

Use restrictions
Access restricted bona-fide research. Submission of research plan and selection of variables; use restricted in time and bound to explicitly stated persons; no access outside the premises of Statistics Denmark and a few other designated institutions (except Eurostat)

Access conditions to previous files
Same access conditions

Acquisition costs
For the moment: Access to the complete data set: DKK 68.000 – for smaller part of the data set the price will be lower. To this comes payment for the use of the bureau facilities, edp-supervising, etc.
The prices or for the moment under revision.

Special user support
Specific on the HBS: Forbrugsundersøgelsen
In general on the use of data in Statistics Denmark: Forskning og metode (Research and methods)

Part title(s)
Results are published in a variety of serials
The most important and regular publication is:
Danmarks Statistik: Forbrugsundersøgelsen xxxx-xxxx (indicating the years of data collection) in the series: Indkomst, forbrug og priser.
To this comes a yearly special article in the same series on a selected main topic.
In the English version of the Statistical Yearbook (can only be found on http://www.dst.dk) some of the tables are published in translated form.

8.2 Labour Force Survey 2001

Title
Arbejdskraftsundersøgelsen

Parallel title
Labour Force Survey

Investigator
Danmarks Statistik

Subject items
The LFS gives first of all detailed information on the populations attachment to the labour market. Furthermore it gives informations on working times and second jobs and the time-length of job-search and the methods used for job-search. Also data on participation in education-activities is given.

Statistics domains
Labour market status, Working conditions, Education activities

Abstract (Study Purpose)
The HBS is a multi purpose survey. Its main purpose is to give a proper description øf the labour market attachment of the Danish population. This description includes a grouping of the population in either employment, unemployment or not in the labour force (e.g. students or pensioners)

Reference period
2001-Q1

Institutional universe
Total population

Geographic universe
Resident population

Type of source
Combination of person survey (for the population 15-74 years old) and registers

Study frequency
Quarterly

Sample Size
16.700 (per quarter)

Sampling method
Stratificied sampling is used to ensure enough unemployed respondents. One third of the sample is drawn from the population registered as unemployed in the latest available version of the unemployment register, while 2/3 is drawn among the persons not registered as unemployed. The information on unemployment according to the register date to 3-9 months prior to the interview.
From each strata the sample is drawn randomly.

Total response rate
73 percent. (in 2001-Q1)

Class of survey
Labour Force Survey

Core Variables
country of birth, citizenship, gender, age, marital status, household relationships, education, continued education, labour status, status in employment, working hours, occupation, industry, main social status, socio-economic classification

Standard Codes
NUTS, ISCED, ISCO-88, NACE R.1

First year of conduction
1994 as a quarterly survey but yearly surveys with similar content has been conducted since 1984.

Conduction history
1994 -: quarterly

Data disseminators
Statistics Denmark, Arbejdskraftsundersøgelsen

Statistical disclosure control methods
Removal of direct identifiers

Type of access
For scientific uses and for Eurostat

Use restrictions
Access restricted to bona-fide research. Submission of research plan and selection of variables; use restricted in time and bound to explicitly stated persons; no access outside the premises of Statistics Denmark and a few other designated institutions (except Eurostat)

Access conditions to previous files
Same access conditions

Acquisition costs
Depends on data (number of variables, number of units, number of time-periods)

Special user support
Specific on the LFS: Arbejdskraftsundersøgelsen
In general on the use of data in Statistics Denmark: Forskning og metode (Research and methods)

Part title(s)
Results are published in a variety of publications. The most important and regular publication is:
- Statistiske efterretninger: Arbejdskraftsundersøgelsen xxxx xx (indicating the year and quarter of data collection) in the series Arbejdsmarkedet.
- In the English version of the Statistical Yearbook (can be found on http://www.dst.dk) one table is published in translated form.
- In Konjunkturoversigten quarterly data is given with explanations in English.
- Eurostat publish the results for all memberstates once a year in: Labour Force Survey
- Furthermore quarterly results are available in the database, http://www.statistikbanken.dk (free). It is planned to make a version in English