**Expert Group Meeting on**
**Setting the Scope of Social Statistics**
United Nations Statistics Division
in collaboration with the Siena Group on Social Statistics
New York, 6-9 May 2003

Use of Registers in Social Statistics in Denmark*

by

Kirsten Wismer **

# Contents

# Use of registers in social statistics in Denmark

*Kirsten Wismer, Statistics Denmark*

## 1 Introduction

Social statistics are concerned with the population and their living conditions. The basic data for these statistics is provided by the classic *demographic statistics:* Total population by age, sex, and geographic area, supplemented by *vital statistics* showing the trends (births, deaths, migrations, etc.). These basic statistics have been expanded to include income distribution and consumption, education, employment, unemployment and labour market policy schemes, wages and labour costs, health, use of public services and receipt of social benefits.

As a result of a traditional division of labour, Statistics Denmark is mostly concerned with the production of statistics based on objective indicators of welfare and living conditions, rather than on perceived levels of welfare or health which are taken care of by other institutions.

A characteristic trait of Danish social statistics is that they are largely based on the administrative registers of other agencies. Data collected directly from the relevant persons and households by means of interview or postal inquiry are limited to the collection of supplementary information in cases where register data are not available or the survey methodology is decided by the contractor, i.e. EUROSTAT.

The statistics covers a broad spectrum of subjects, but together they constitute a single coherent system. Most of the statistics are produced on a yearly basis, some of them more frequently. The use of a single, coherent system of statistics means that, since 1981, even Denmark's population and housing census has been based solely on administrative registers, with no questionnaires to citizens. This means that new and detailed census statistics are available every year.

## 2 Administrative data as statistical data

In order to use administrative registers in official statistics certain requirements have to be fulfilled. These requirements concerns coverage of the data in terms of content and some qualitative aspects, but it is impossible to give very precise definitions. In the following some requirements are set out if you are creating a coherent, comprehensive system of personal statistics based on registers, i.e. a system of the kind developed in Denmark.

### 2.1 Precise definition of units

For statistical use it is important that the units in the administrative registers are well defined. For a 'natural' unit such as persons this is not a problem. Also the concept of a dwelling is fairly easy to deal with while a household is more complicated. By defining a household as a dwelling household is it possible to use a register with the dwelling as a unit.

Other types of units are much more complicated as for example business units. A firm or a workplace is an abstract concept. Its existence can only be defined and established in accordance with very precise rules.

## 2.2  Total coverage

Registers should, ideally, contain information on all units in the country of the types specified, i.e. they must be comprehensive. Incomplete coverage will often be systematic, for example, if certain municipalities are not covered, the statistics produced on such a basis will be affected by bias which it may be difficult to compensate for.

## 2.3  Identifiers

Identifiers play a very considerable role in the maintenance of administrative registers and in their statistical use. For statistics it is particularly important that it should be possible to employ the common identifiers used in many registers in order to link information from various sources on one unit. Since administrative registers are not set up for statistical purposes and therefore do not usually contain the desired data combinations in directly accessible form, an important procedure in the production of almost any register-based statistics is to link data from different registers by means of common keys.

The most important key in this connection is of course the person number (PIN-code). The street address and the workplace number are also essential in determining relationships between persons, dwellings and workplaces.

## 2.4  Time references

The time dimension plays a very special role in statistics revealing patterns and trends in society, and in all areas of coverage it is necessary to be able to make comparisons over time. In statistical publications or databases the time dimension is, broadly speaking, ever present. It is therefore vital for statistical usability that reliable information on various dates be contained in registers.

This applies first and foremost to dates of changes or events. Among the main events of interest are the 'birth' and 'death' of units, but it is also important to date other changes affecting units. What we are concerned with here is the real point in time at which an event took place bringing about a change in an item of data.

In addition to dates of event or changes there is a need for registration dates, as an indication of when the data value in question was entered in the register. This is almost as important as date of event when the register has to function as a statistical model of the reality the statistic is to describe.

## 2.5  Stability

An important property of a statistic is that it can describe a process over time. Therefore it is of great importance, that the concepts in the administrative registers remain constant. Otherwise major problems can arise in securing comparable figures from one period to the next.

In some cases it may be possible to adjust for changes with greater or lesser precision. The income statistics is an example of this. Through the years there have been many changes in the tax-laws, which have changed the content in the administrative data in the tax-files. Even so it has been possible to maintain the register based income statistics through a very careful adjustment from administrative data to the statistical data.

## 2.6   Quality

The quality requirements imposed by statistical use coincide to some extent with the requirements which must also be met in serving the primary purpose of the registers.

The first quality requirement to be met by basic data is that they must be *relevant* to the statistic. This means that they must relate to the concepts used in the statistics. In favourable cases the statistician may exert an influence on the content of the administrative data. Some times it is possible to let different data sources supplement one another through linkage, so that the desired information is provided by comparing a number of data from different registers.

It is important to notice that a statistical system which for a large extent is based on administrative sources may end up giving a picture of the world seen through the eyes of those authorities. Concepts or persons who do not exist in the frame of the tax authorities or the social workers may not show up in the statistic. An example from the Danish system is information concerning homeless people.

To exclude this risk it is important that there exists a basic register, whose task is to record all units without reference to any specific administrative purposes. In section 3 is the description of the 3 basic registers in the Danish statistical system.

The next requirement concerns *reliability*. There must be a high degree of certainty that the recorded data faithfully reflects the circumstances investigated.

Finally it is important that data are recorded with a degree of *precision* suited to the needs of the statistics. This is important even when information is not to be used in statistical reports at the recorded level. In many cases groups are formed on the basis of variables derived from discrete items of data.

## 2.7   Changes in the administrative data

As has been said, major problems may be posed for a statistic when legislative or regulatory changes result in alterations to the data content of the administrative registers. On one hand, it may be difficult or impossible to assess the long-term trend in a particular magnitude if different definitions are used in the base material. On the other hand, problems may arise in deciding what changes in data values are to be viewed as reflecting actual events and what changes merely represent new concepts or definitions.

The statistical consequences for data altered because of new legislation depend on what type of statistic is involved. If it is in fact a statistic for the monitoring of legislation the statistic of course merely has to go along with and adopt the concepts of the new legislation.

If we are concerned with a more general statistic it is not acceptable that it should be impossible to compare the statistical concepts before and after the legislative changes. In this case an attempt must be made to estimate the significance of the changes – at least on an aggregated level to ensure that time series can be created with chaining.

## 3   The basic registers used in Social Statistics

### 3.1   The Central Population Register

With the Population Registration Act of 10 June 1968 a decisive reform was carried through, aimed at simplifying and increasing the effectiveness of the entire system of personal registration using electronic

processing technology. Alongside the municipal population registers, which had existed since 1924, a nation-wide register of the Danish population - the Central Population Register (CPR) - was set up on magnetic tape. The Central Population Register was built by adding the information from the population registers in the municipalities. The updating of the Central Population Register was done by the local registers.

A crucial part of the reform was the introduction of a fixed identification number for each individual - the person number. This number was considered to be a practical necessity for the maintenance of the Central Population Register. The same number was to be introduced throughout public administration and would thus replace the many numbering systems which had previously been used in various administrations.

In conjunction with the establishment of the CPR a number of other standards and codes were introduced, for example, municipality codes, parish codes, occupation codes and nationality codes. Most crucial was the address code, which identified the individual address; this formed a crucial element in the CPR-system.

The motivation for setting up the CPR was, first and foremost, to avoid double recording and the extra consumption of resources which that would engender. Contributory considerations were the prospect of a tax reform with the introduction of a tax deduction at source (implemented in 1970) which would be difficult to carry through without a reliable system for identification of the country's citizens.

The content of the Central Population Register is information on name, address, sex, date of birth, marital status, place of birth and citizenship for every person living in the country. The updating of the register is done by reporting births, deaths, migrations and so on by the local registers.

## 3.2    The Central Business Register

The register, which is administrated by the tax authorities, contains information on self-employed people and employers who are entitled or obliged under other legislation to register with public authorities or institutions. In the register the employer was identified by a CVR- number (a number that was given by the tax authorities to use in the settling of the tax at source) or - for the self-employed - the person number. For the units is information on industry and type of ownership.

In order to link an employed person and his workplace or establishment a code for the workplace was introduced in 1979. In case of a private employer with only one establishment, the workplace and the establishment are identical. Employers with more than one workplace are every year asked by Statistics Denmark to identify each workplace by a code. Each workplace will thereafter be given a code for industry.

The introduction of the workplace in the registers opened the possibility of obtaining commuting statistics.

## 3.3    The Register of Buildings and Dwellings

In 1977, as a result of the decision not to conduct a conventional census of population and housing in 1975 and for the purposes of property valuation, information was collected for a Register of Buildings and Dwellings (BBR). The register was an extension of the national computerised Property Register, which was set up in 1972. The register is maintained by the municipalities by reports on alternations, new buildings and demolition, supplied in conjunction with planning applications processed by the municipalities. The register is organised hierarchically in three levels: property, building within the

property, housing unit or other unit within the building. The register has information on age, size, installations and use of buildings and dwellings.

The housing units or other units are given addresses in a format that is the same as in the Central Population Register (municipality code, street code, number in the street, floor number, location on the floor). It is through the address coded in this way that a linkage can be made between the housing characteristics in the BBR and the personal details in the CPR. Continuous communication between the two basic registers ensures that addresses are identical so that the information in the registers can be interlinked.

## 4    Social statistics based on registers

Over the past 30 years there has been an extensive use of administrative registers in the production of statistics in Statistics Denmark. The result is that a great part of social statistics today is based on administrative registers. This development was foreseen already in the sixties when the Act on Statistics Denmark was framed. The Act contains extensive provisions allowing it to collect information from public authorities.

**Table 1** gives an overview of the social statistics in Denmark. It shows the content of the register statistics, which are solely annual surveys.  The exception from this is the population statistics and unemployment statistics, which are monthly, and statistics on labour market policy and crime, which are quarterly. The main part of the system is based on administrative records, where the responsibility for the registers lies at outside Statistics Denmark. In some areas there does not exist one administrative source. Therefore Statistics Denmark has established statistical registers based on information collected from institutions, local authorities and business. The common part for all registers in the system is, that it contains information of all individuals in the population and that each individual is identified with the personal unifier.

However the register statistics do not cover all social statistics in Denmark. Firstly, in some areas there is a need for more frequent statistics than annual, and secondly, the cooperation within EU implies the need for some household surveys. But also for these statistics we use information from the register statistics if possible.

Below is a short description of a couple of the social statistics based on administrative sources. Further information can be found in the Declarations of Content available on www.dst.dk.

**Table 1. An overview of social statistics in Denmark**

| | Based on administrative sources | Business survey | Household/person survey |
|---|---|---|---|
| **Register-based system** | | | |
| Population and education | Population (p)<br>Census (p, e)<br>Family and households (p)<br>Migration (p) | Education (p)<br>Training (p) | |
| Labour market | Labour force (p, e)<br>Unemployment (p)<br>Labour market policy (p)<br>Wages and labour cost – public sector (p, e) | Labour market policy (p)<br>Wages and labour costs - private sector (p, e) | |
| Social and health | Social protection (p)<br>Pensions (p)<br>Use of hospitals (p)<br>Criminal statistics (p)<br>Traffic incidents (p) | | |
| Income and Registers | Income statistics (p) | | |
| **Independent surveys** | | | |
| Labour market | Employment indicator (e)<br>Wage sum (e)<br>Strikes | Wage index – private (e)<br>Wage index – public (p, e) | Labour force survey (p) |
| Social and health | Staff in social institutions (p, e)<br><br>Day-care (p) | Social production | |
| Income and Registers | | | EU-SILC (p) |
| Consumer and Interview Service | | | Consumer expectation survey |
| **Harmonized statistics** | | | |
| Labour market | | Labour Time Account | |

Note: p: person number is applied in the survey
e: enterprise identifier is applied in the survey

## 4.1   Population statistics

The population register is based on data from the Central Person Register. It receives annual data of the total population and weekly data with information of demographic events such as removals, emi-/immigration, birth and death.

The population statistics are based on the resident population in Denmark, and statistics are published quarterly and annually for local areas. The statistics include statistics on births, deaths, marriages, divorces, surviving spouses, registered partnerships, relocations, emigrants and immigrants, naturalisations and adoptions.

The size of the Danish population is calculated on 1 January each year. Besides the publication of these statistics the population register is used to establish populations in the rest of the statistics in the register system. Furthermore it is used to draw samples for household surveys.

## 4.2 Statistics on immigrants and their descendants

The statistics on immigrants and their descendants were introduced in 1991. The concepts were introduced to make it possible to give information on population with foreign background, incl. people who have attained Danish citizenship. The statistics are solely based on the Central Population Register.

An immigrant is defined as a person born abroad whose parents are either foreign citizens or were both born abroad. A descendant is defined as a person born in Denmark whose parents are either immigrants or descendants with foreign citizenship.

Statistics are published annually and contain information on country of origin, citizenship, gender and place of birth and residence. These statistics are used as important background information in most of social statistics.

## 4.3 Register-based statistics on labour force

The register-based statistics on the labour force (RAS) are annual statistics, which are based on a number of sources. Common for all sources is that they are all statistical registers, which has been formed in other parts of Statistics Denmark.

The primary sources are *the population statistics* to define the population of the statistics and *different tax registers*. As described above the population register is established in the population statistics. The tax registers are transformed from administrative records to statistical information in a unit in the business statistics. This division of labour ensures the same database for employment statistics in local units, which is performed in business statistics, and employment of the population, which is performed in social statistics.

Other sources used in the register-based statistics on the labour force are:
- wage statistics
- the Central Business Register
- statistics on labour market policy
- statistics on education
- statistics on unemployment
- the employment classification module
- statistics on benefits during sickness or in connection with child-birth
- statistics on social benefits
- statistics on social pensions

The statistics gives information on the population's attachment to the labour market at a given point in time at the end of November. According to the ILO guidelines the population is divided in three groups: persons in employment, persons in unemployment and persons outside the labour force. Employed persons are either employees, self-employed or assisting spouses.

The statistics are compiled in two stages. In the first stage, the gross stock of self-employed, assisting spouses, employees, unemployed and persons on early retirement is compiled. A person is for example to be categorised as an employee if the person has a certain degree of attachment to the labour market and if the person is employed at the time of reference. A person is defined as unemployed if the person is unemployed at the time of reference.

A person can be included in more than one gross stock, and it is therefore necessary to select the most important labour market status for each person. This has led to the following ranking of the various gross stock:

1. Unemployed
2. Early retirement pensioners
3. Employers
4. Employees with jobs at full-time equivalence
5. VAT-registered persons
6. Self-employed insured against unemployment
7. Employees with jobs at part-time equivalences, or duration of employment not stated
8. Other self-employed
9. Assisting spouses

Persons outside the labour force consist of 9 groups with the following prioritisation:
1. Recipients of early retirement
2. Participants in labour market education schemes
3. Recipients of benefits during sickness or child-birth
4. Recipients of rehabilitation or cash benefits
5. Children and young people
6. Recipients of old age pension
7. Recipients of civil servants earned pensions
8. Persons in education
9. Others outside the labour force

The labour force statistics contain a number of variables such as gender, age, immigrants and their descendants, municipality of residence, municipality of work, nationality, education, industry, socio-economic status, sector, ownership, salary, fund, participation in labour market schemes, degree of unemployment in the year, insurance category and the number of hours worked.

The drawback of such comprehensive statistics is that it has a rather long production time. In recent years it has been published about 15 months after the end of the reference year. The latest statistics published in March 2003 refers to the population at January 1 2002 with the attachment to the labour market in November 2001.

## 5    Reorganisation of the register system

The current Danish register system for statistics on persons is to a very high extent influenced by the stepwise development during the last thirty years. The number of data has increased heavily and the use of general identifications, mainly the PIN-code, has led to plenty of possibilities for producing statistics based on varying combinations of variables. Researchers and others are very much aware of that situation and they are very keen on utilizing the possibilities of analyzing specific topics in the different fields of living conditions.

By now data are organised in registers as physical units of the statistical system and still more registers have been created completely or partly containing variables from other registers. For different reasons it had to be so, but the system as a whole is now so big that hardly anybody can cope with it, and the continuing extensions that are expected will contribute to further complications. Therefore reorganisation of the register system has been discussed in Statistics Denmark and is now on its way.

The results of the reorganisation are expected to be:
- Harmonised definitions of concepts used in different fields of statistics
- Clarified responsibility for creation, maintenance and documentation of each variable
- Simplified and more transparent documentation
- Improved guarantee for optimal choice of variables to be used in the individual case
- A more efficient operation of the system
- A higher degree of flexibility by means of knowledge sharing achieved by closer contact between collection and deduction of data and the final use of the information
- Less space required for storage of data

These improvements will contribute to ensure the quality of the statistics and a more optimal use of the resources.

**References:**

Finn Spieker, "A statistical system of registers based on logical data connections", Statistics Denmark, November 11, 2002.

Declaration of contents, www.dst.dk

Lars Borchsenius, "From a conventional to a register-based census of population", Statistics Denmark, November 2000.

Eurostat and Danmarks Statistik, "Statistics on Persons in Denmark, A register-based statistical system" (1995).