

METHODOLOGICAL PROBLEMS CONNECTED  
WITH A SOCIO-DEMOGRAPHIC STATISTICAL SYSTEM  
BASED ON ADMINISTRATIVE RECORDS

Lars Thygesen

Danmarks Statistik  
Denmark

1. INTRODUCTION

The perspectives of utilizing administrative records for statistical purposes have been discussed intensely among statisticians for many years, especially at the international level. This was also the case at the 43rd session of ISI in 1981 in Buenos Aires. These discussions have revealed a widespread opinion that the necessary relevance, continuity, and reliability of statistics can only be obtained by applying traditional statistical data collection methods.

Many Danish statisticians take a more optimistic view on register-based statistics. They also point out that quality problems concerning relevance, continuity and reliability concern all statistics: Traditional statistics may be affected by similar errors that seem less serious because statisticians have become used to them.

When founding Danmarks Statistik (the central statistical office) in 1966, the Danish legislature foresaw a strong movement towards register-based statistics, the aim of this evolution being to reduce the response burden on private citizens as well as that on the business sector. A tight budget situation for Danmarks Statistik has contributed to speeding up the transition to register-statistics. Since the early seventies a full register-statistical system has been a goal in the strategic planning of Danmarks Statistik.

This choice of strategy depends on the fact that administrative registers in the Nordic countries and especially in Denmark are well developed and suited for statistical use. The existence of the registers of course makes it difficult to employ traditional statistical data collection: It is difficult to persuade respondents to give information that they have already reported to some other public authority.

In some branches of statistics the overall register-strategy implies the use of data collection methods that would not be regarded as optimal if each branch was seen in isolation. In other cases, it may be necessary to abandon collection of useful information that could have been obtained at low marginal cost by traditional methods. Thus, the choice of strategy is a real one, and it is decisive for a number of other decisions.

The Danish position has been expressed most clearly by the decision to base the 1981 Population and Housing Census exclusively on data from administrative records.

The development of register-statistics raises a number of new methodological problems, many of which have not been sufficiently explored. It is the aim of this paper to identify and discuss some of the most important problems.

As a necessary background for this discussion, sections 2 through 4 review the administrative and legal preconditions of the Danish development. Section 5 outlines the present statistical system in Denmark. Section 6 presents and discusses the methodological questions.

## 2. ORGANIZATION OF THE STATISTICS PRODUCTION IN DENMARK

The official Danish statistics are highly centralized. The institution responsible for the production of the largest part of the general statistics on demographic, social and economic conditions etc. and for the general co-ordination of such statistics is "Danmarks Statistik".

This institution was in its present form created by an Act of Parliament of 8th June 1966. It is an independent institution under the supervision of a Board consisting of the National Statistician as chairman and six other members familiar with social and economic conditions. The board is appointed by the Minister for Economic Affairs.

The Board decides on Danmarks Statistik's working plan and lays down rules for the co-ordination problems.

Furthermore, the Board decides the extent to which and the way in which information under the provisions of the Act must be collected.

The Board has in these matters the final authority, whereas other matters - especially those concerning budget and personnel - must be submitted to the Minister for Economic Affairs.

Of special importance as regards the theme of this paper are two basic principles of very long standing. The first principle is that Danmarks Statistik - now according to the Act - is entitled to collect administrative data from public authorities for the production of statistics, and to organize the necessary co-operation with the various administrative registers.

The second principle is that data collected for statistical purposes and referring to an identifiable person or enterprise may not be passed on to any other government department or private person. This "principle of non-disclosure" has been strictly observed during a very long period of years. The only exception is that the Act on a Central Register of Enterprises and Establishments specifically authorizes Danmarks Statistik to furnish this register - which is administered by Danmarks Statistik - with some basic data, such as

identification data on the units and data permitting their classification by kind of economic activity and by size according to total employment. Such basic data may according to the act be passed on from the central register to government departments and to a certain degree to private persons.

### 3. DANISH PUBLIC ADMINISTRATION AND REGISTRATION

Public administration in Denmark relies rather heavily on recorded information concerning the individual objects of their administration: Citizens, business enterprises, buildings etc. This is a tradition that goes far back in time. For the purpose of this paper the focus will be on describing registration of persons. Registration of business units and buildings and dwellings will also be briefly described.

#### 3.1. Registration of persons

Back in 1924, the Danish parliament passed an Act on Local Population Registers, which laid down that in that same year all Danish municipalities should each establish a local population register, ie. a file containing information about all persons resident in the municipality district, irrespective of their being temporarily absent.

The local population registers were to contain information for personal identification, such as occupation and name, date and place of birth. Apart from this the most important items of information registered were place of residence, family circumstances and citizenship.

The municipalities were to keep the files continually up to date by means of information about births, deaths, marriages, divorces, etc., received from various authorities. The individual persons themselves became liable to give notification of address changes direct to the local population registers.

In 1968 an important reform of the local population register system was implemented, the purpose being to render the total public registration of the population more efficient by means of electronic data processing. In addition to the local population registers, which continued to exist, the Central Population Register, CPR, was founded as a national magnetic-tape register of the Danish population, a major benefit being that the CPR and the local population registers can be updated by one integrated administrative process.

An essential part of the reform was the introduction of a permanent and unique identification number for each individual person: the Person Number. This number was considered a practical necessity for the operation of the Central Population Register. Moreover, the number was to be introduced into all sectors of public administration, replacing the many different number systems which had previously been employed by the various branches of administration.

The establishment of the CPR was first and foremost brought about by the wish to avoid multiple registration and thus to save resources. Another con-

tributing factor was the prospect of a tax reform introducing the system of taxation at source.

The first direct consequence for the statistical work was that the annual vital statistics were reorganized in the period 1970-1973, so that by far the main part of their primary data are derived from the CPR. However, the CPR data concerning births and deaths are still supplemented by medical data from the health authorities. Since 1970 it has been possible - based alone on the CPR register - to compile annual population statistics with distribution by sex, age and marital status for highly disaggregated regions.

In relation to the statistical use, some of the important features of the CPR are:

- the register is used by almost every public authority, which helps to improve the data quality,
- the person number is a secure link to other person registers,
- family ties are indicated in the CPR, making automatic linking of family members possible,
- households are indicated by exact address designations, which also makes possible the linking with data from the Central Register of Buildings and Dwellings mentioned in section 3.3.

During the late sixties and the early seventies the use of edp in the Danish public administration grew very fast. This applies to central as well as to local government. Among the larger fields in which edp was introduced can be mentioned the Danish withholding-tax scheme (1970) and several social security schemes, such as social pensions, children's allowance etc. The registers established to serve these schemes are valuable sources of statistical information.

To a large extent the data processing in the technical sense is performed by two large edp bureaus owned by the state and the municipalities. However, some government agencies, including Danmarks Statistik, have their own edp installations.

### 3.2. Registration of businesses

In the late 1950's, Denmark's central statistical institution decided to create a business register, which was to serve statistical purposes only. The reason for the decision was the inadequate potentials for selecting representative samples for different fields of economic statistics.

The basis for the register was to be a business census conducted in 1958 in all industries except agriculture and fisheries. The register was to be updated by means of information from address books and other available sources. It became possible to start utilizing the register in a few fields in the early 1960's. However, a complete and comprehensive register system with reasonably fast updating could not be achieved by that method.

This situation was radically altered when value added tax was introduced in Denmark in 1967. In that connection the customs authorities set up a

register of all business units liable to settle VAT in all sectors of the economy, except primarily the health services and passenger transportation. The registered data were made available to Danmarks Statistik's business register, and this material was utilized for the production of statistics concerning sales, purchases, etc., in the various industries.

Additional improvements of the business register were achieved when it became possible from around 1970 to obtain further data on employers from the withholding-tax administration that was created in 1970 as a part of a tax-at-source reform.

As business registration in the statistical field continued to improve, there was a more and more pronounced wish for setting up an administratively oriented business register containing a few basic data types, such as identification code number, name and address, kind of activity performed and size category of the individual business units.

Such a register was founded under a special law in 1975, and called the Central Register of Enterprises and Establishments. The operation of the register was entrusted to Danmarks Statistik, so that in this particular field the Danish central statistical office is in charge of an administrative function.

The Central Register of Enterprises and Establishments provides the basis for all of Danmarks Statistik's business surveys, either in the phases of sample selection and addressing of questionnaires or for the matching of data from other administrative sources, cf. Annex 1, item 1. With the availability of statistics produced in this way, it has not been considered necessary to conduct a traditional business census since 1958.

As its name implies, the register contains business units at enterprise level (legal units) and at establishment level (local units). For owners of enterprises the person numbers, too, are entered on the register.

### 3.3. Registration of buildings and dwellings

In the early 1970's there was widespread demand for setting up a nationwide register of buildings and dwellings for the benefit of local government planning and for use in the general public assessment of real property, so far carried out every four years for taxation purposes. It was considered how such a register could be organized with a view to also serving statistical purposes, notably in the fields of housing statistics (including housing censuses) and construction statistics. In that same period it was discussed whether the population and housing census for the mid-1970's should be conducted by traditional questionnaire-based methods or possibly by alternative methods. At the time, the population registration was already quite extensive, and it was therefore only natural that the matter of setting up a register of buildings and dwellings should be included in the overall considerations. The outcome was that in 1974 the government decided that a Central Register of Buildings and Dwellings ("Bygnings- og boligregistret", abbreviated BBR) should be established, and that there should be no population and housing census for the mid-1970's conducted by traditional methods.

A special law concerning the setting up of the BBR under the Ministry of Housing was enacted in 1976. The original data basis of the BBR was obtained in connection with the general public assessment of real property for 1977, and the updating is carried out by the local government authorities in connection with their supervision of new building projects.

Following various pilot projects in 1977-1979, annual housing statistics and quarterly construction statistics have been compiled on the basis of the BBR since 1st January 1980.

Special effort has been made to ensure that the address designations (the way addresses are written and spelt) in the CPR and in the BBR are exactly alike, so that via the address designations the dwelling data of the BBR can be matched with the person data of the CPR. Experience has now revealed that in Denmark this solution leads to a satisfactory degree of congruence between the address designations of the two register types. Thus for the dwelling stock statistics and the housing census at 1st January 1982 the housing situation of only 1.2 per cent of the population could not be accounted for. The corresponding figure was 4.0 per cent for 1st July 1979.

#### 4. PRIVACY CONSIDERATIONS

One principal precondition for the development of register-based statistics is that this development is accepted by the general public and by the legislative power. In this respect the Danish situation is and has been favorable.

Like in many other countries there has in Denmark been public concern about the use of edp register techniques by public authorities. Contrary to what one might have expected, the debate did not catch on when the big, administrative edp systems were established around 1970. Not until 1977 did these questions to a larger degree attract public attention.

When they did it was in connection with, among other things, the introduction of new statistics on recipients of social benefits, the need of which was created by a new Act on Social Assistance. Danmarks Statistik decided that the statistics should be based on data on the individual recipients of such benefits, identified by their person numbers. This technique was chosen mainly to reduce the respondent burden on the local governments to a minimum.

A few municipalities claimed their clients' right to anonymity and contested the right of Danmarks Statistik to demand such information, not accepting that the above-mentioned statutory provisions in the Act on Danmarks Statistik would apply in this case. It was argued that the use of person numbers, not least for statistical purposes, was dangerous and irrelevant. Eventually the Danish Supreme Court decided that Danmarks Statistik without any doubt had the right to demand - and the local governments the duty to pass on to Danmarks Statistik - identified information on recipients of social benefits for the statistical purposes in question.

In 1978, the Danish Parliament enacted two bills on registers. One of them, entitled the Public Authorities' Registers Act, which became effective January 1, 1979, clearly stresses the rules of nondisclosure of statistical data that were already in effect as well as the rules of accessibility of register data. Moreover, it ordered the establishment of a Data Surveillance Authority (DSA), to supervise the registers and the regulations concerning their operation with a special emphasis on safety measures for the protection of information. The final competence in decisions regarding the setting up, etc., of registers rests with the political authorities (ministries and local government councils), but before a decision is made the DSA must be heard. The DSA has the power to exercise control by direct inspection of the installations where the registers are kept. The DSA must be notified in each case of matching of registers and the DSA may establish the conditions of how such matchings shall be executed.

Although the primary aim of the Act is to regulate the use of registers in public administration, the regulation also applies to registers that are used only for statistical and scientific purposes. However, the rules applying to such registers are different. Thus the linking of information from different registers can be administered without notification to the DSA when the purpose is strictly statistical, and the right of a person to know what the registers hold on him is not extended to statistical registers. The reason is that statistics cannot threaten privacy, because information about individuals cannot be identified from the statistical tables, and because individual information on statistical registers may not be passed on to others, nor be used for administrative purposes.

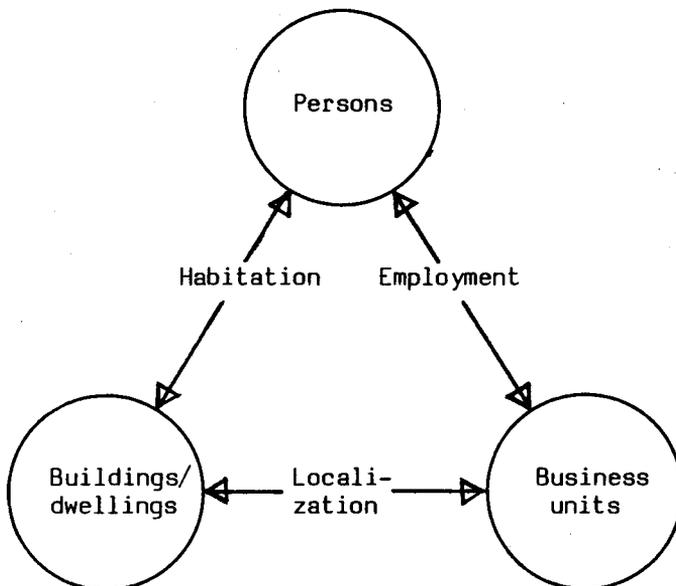
## 5. PRINCIPLES OF THE DANISH STATISTICAL SYSTEM

The Danish system of register-based statistics has evolved gradually since 1968. Some of the important events and decisions are shown in Annex 1. This section contains a brief description of the socio-demographic part of the present system.

### 5.1 A simple model

The central objects of the socio-demographic statistics are the persons. We need to know the relations of the persons to the dwellings in which they live, and to the business units where they work (workplaces) if they are employed. Thus, there are three kinds of objects which are interrelated as shown in fig. 1 below.

FIG. 1: A simple model



Another important relationship in the model which is not shown in the figure is that between family members.

Each type of objects has a number attributes e.g. age, occupation, number of rooms in the dwelling, that may have a bearing on the social situation of the persons.

## 5.2 The data model

The aim of the statistical system is to create a description of the socio-demographic situation of the persons and changes in that situation. The description should enable the carrying out many kinds of socio-demographic analysis. We should try to get data on the existence and the birth/death of objects, relations between objects, important attributes of objects, and changes affecting relations and attributes.

In Denmark it is possible to extract most data necessary to establish this model from the administrative registers mentioned in chapter 4. The statistical system has the same structure as shown in fig. 1. Each corner of the triangle is a number of more or less independent statistical registers, each containing the data necessary for one field of statistics, e.g. personal income statistics. There are by now some 35 statistical registers. They can be linked together by means of the unique and common identifiers of each type of object, most important of which is the person number. The relations between the different kinds of objects shown in fig. 1 are expressed in terms of the same identifiers.

A special kind of registers are called Systematized Data Modules. These are not aimed at one branch of statistics, but they contain data for general use in several fields. One example is the Employment Classification Module, furnishing employment data for a wide range of person statistics. One advantage of this arrangement - apart from its being resource efficient - is that comparability is secured between different branches of statistics, and when different groupings are used the differences can be reconciled. This makes for a suitable balance between the demands for standardization and flexibility.

In many cases, one statistical register is based on data from a number of administrative registers, because the data combinations used by the administration are not relevant to users of statistics. Consequently exact matching by means of person numbers etc. becomes a vital part of updating procedures.

Experience shows that even the most reliable administrative data must be submitted to statistical data processing before entering the statistical registers. The reasons are:

- the time reference of register data must be adjusted because of varying delay in the updating of administrative systems,
- data must be controlled and validated (especially when the information in question is not used systematically by the administrative authority),
- it is necessary to compare and adjust data on the same subjects, stemming from different sources - sometimes a very complicated task.

### 5.3. Use of statistical registers

The system is geared to two main types of uses: on the one hand the utilization of a single register in specific fields of statistics, and on the other hand, ad hoc utilizations of two or more registers or systematized data modules by matching data selected to suit the individual projects.

## 6. METHODOLOGICAL PROBLEMS

The purpose of the rather extensive description in the preceding chapters of Danish experience in the register-statistical field has been to serve as a background for the discussion of the methodological problems.

The setting up of the general statistical system involves a large number of methodological problems and of course some of them have by now found solutions - more or less perfect. In this chapter we shall look into problems that have not been sufficiently explored.

### 6.1. "Integrated data collection"

By the term "integrated data collection" we shall understand collection through the administrative systems of data that are used for statistical purposes only.

Even in a country with a strong tradition for registration it is very unlikely if not inconceivable that all data necessary for the general statistical system should be contained in the administrative records. In such cases it may be a satisfactory solution to collect the supplementary data by sample surveys. But sometimes the supplementary data must be made available with a total population coverage, e.g. to allow compilation of tables with a high degree of geographical disaggregation. Is it possible then to make the administrative authorities extend their records with statistical data that are not needed for the purposes of the administrative authorities themselves? And if this is possible, will the resulting data have a sufficient degree of reliability?

Of course the answer to the first question depends very much on the attitude of the administrative authorities and the legal rules in the country. In Denmark both of these conditions have been rather favorable, and co-operation has been possible in a number of cases as long as the demands of Danmarks Statistik have been kept at a modest level.

The question whether integrated data collection may lead to satisfactory results is a more difficult one. The problem is of course that the authority responsible for the collection tends to see the extra statistical data as less important and as a burden. Consequently, they may not try hard enough to avoid non-response among the citizens providing the information, and they will take little interest in setting up data control procedures etc.

Two cases of Danish experience will be briefly commented upon:

- (1) Collection of information on occupation, and
- (2) "The workplace project".

#### 6.1.1. Information on occupation

Information on occupation of the economically active population is a central datum in a socio-economic statistical system. In Denmark this information is collected especially for Danmarks Statistik through the annual tax return form from every tax-payer.

The municipal tax authorities are responsible for recording the citizens' messages of shift in occupation and for passing on the information to Danmarks Statistik. It has become obvious that a number of municipalities are not meeting this obligation as they report very few messages of shift. Danmarks Statistik has tried to improve the situation by information campaigns, but the results have not been encouraging. Another problem is that many citizens do not care whether the registration of their occupation is correct, as this will have no administrative consequences.

These problems have made it necessary to use as a supplementary source all available information on occupation from other registers, covering a smaller or larger proportion of the total population.

Still, the information on occupation remains a somewhat weak link in the total system.

### 6.1.2. The workplace project

For the 1976 register-based population census a system was developed for the kind-of-activity distribution of the economically active population. The method involved the matching of data on enterprises from the Central Register of Enterprises and Establishments with data from a special tax register of employers' payments of wages and salaries to their employees (the "Salary Information Register"). The system worked in practice, but it was a problem that the link between employees and their business units was supplied at enterprise level and not at establishment level, so that the kind-of-activity distribution of the population could not be in accordance with statistical needs.

The objective of the workplace project, from and including 1979, was to collect the missing information on the relation between employed persons and the physical workplace, i.e. the establishment. The project also made it possible to compile annual commuting statistics and structural business statistics.

Under the project, systematic registration is effected of all local places of work (establishments) and their employees. The information concerning the existing places of work is by and large collected by traditional methods (an annual enquiry among employers).

In the public sector, the link between workplaces and employees is derived from the computerized pay systems for the central and local government agencies.

As regards the private employers the link is the information passed on by employers to the tax authorities on pay slips, cumulated in the "Salary Information Register" mentioned earlier. For that purpose the items of the slips have been extended by a box space to be filled in with a workplace code number (supplied by Danmarks Statistik) for employers with more than one establishment.

Danmarks Statistik checks the link data provided by the employers and takes direct contact to employers for verification.

Work on the project started in 1980, and it has been realized that the error checking and verification task is a "manual" process involving a considerable amount of work (three qualified man-years in 1981). But the principle of integrated data collection has been a success in this case. This is ascribed to the close co-operation between Danmarks Statistik and the "respondents".

### 6.2. Missing data

Register-based statistics are harassed by the problem of non-response well-known from traditional statistics: It is not possible to collect all relevant information on all objects in the population. In register-based statistics missing data occur in varying degrees. This is because one datum in the statistical register, e.g. employment status, may be compiled from elementary data from several sources.

Data on one person may be missing in all sources or in one or more of the sources. The latter situation may be called "partly missing data".

The non-response problem might be handled by imputing the missing data employing certain rules and utilizing the accessible information on the object. Or one could take the simpler approach of many sample surveys and assume the same distribution of non-respondents as that of respondents when blowing up figures to total population size. The latter approach would create problems in a coherent system, where the different statistical registers often are to be linked together to produce one statistical task. The blowing up would have to take place in connection with every match of registers.

The missing data will seldom be totally missing. For example, the information of the exact occupation of a person may be missing but we may know the person's workplace (including kind-of-activity), amount of work rendered and salary earned. This partial description of the person's employment situation should not be lost. Every fragment of relevant information at hand should be utilized. At the same time, one uniform scale should be employed for measuring each relevant attribute in the statistical register.

The problem is complex, and we know far too little about the consequences of employing different methods of solution. Some possible ways of filling in the gaps can be illustrated by way of examples.

- a. Compiling labour force statistics, we need to know whether a person is employed at a certain date (end of November). This information ought to be contained in the Salary Information Register mentioned in subsection 6.1.2, but in some cases it is missing. Here we may utilize another datum in the register, indicating the total number of working hours during the year. If a person has been employed during 90 per cent of the normal working hours, it is very likely that he was also employed at the end of November. The opposite conclusion may be drawn if the amount of work was only 10 per cent. The attribute "employed at the end of November" may then be assigned to the person with probabilities 90 per cent and 10 per cent, respectively, in the two cases.

This way of assigning missing data values has in fact been used. The consequences have not been analysed so far but they will of course depend of the use to which data are put. A one-dimensional table distributing persons into the categories "employed/not employed" will probably give a reliable picture, but the problems grow as more dimensions are included.

- b. As shown by the next example, the nature of partial non-response will often be more complex. In labour force statistics the distribution of employed persons by occupation is essential. If the information on occupation is missing, a number of other characteristics of the person may indicate probability of certain occupations. These characteristics may include kind-of-activity of the workplace, employee/not employee, salary seen in relation to number of working hours. Establishing a probabilistic procedure for assigning occupation in this case would require detailed analysis of the co-variation between occupation and the indicators in question.

The two examples are not different in principle but of course in complexity. The methodological problem stated here and its solution bears resemblance to what is known in the literature as Statistical Matching. Hopefully, the growing experiences of statistical matching will prove useful.

### 6.3. Supplementary data collection: Estimation and control

As mentioned in the introduction Danmarks Statistik has decided to follow a register-strategy in the production of statistics. Even so, traditional data collection by means of interviews or postal inquiries have an important role to play. Some indispensable statistics, e.g. on attitude data, may be compiled in this way only. But there will also be a need to collect data similar to those available from the registers. The purpose of such "duplicate" data collection is partly to create a basis for estimation procedures in register-based statistics, partly current quality control.

The estimation procedure must provide the best translation possible from administrative data to the variables and scales needed in the statistical system. The choice between alternative procedures of translation has so far been mostly speculative: Based on knowledge or assumptions about the credibility of the sources, and based on analyses of accordance and differences between conceptually similar data in the different sources.

The shortcoming of this speculative approach is that the outcome of alternative sets of rules is not tested against the "real world". There is a need to compare the content of the source registers, object by object, with measurement of the highest obtainable quality - that is sample surveys with a reasonable sample size, carefully prepared and conducted by interview. Of course even this kind of measurement is afflicted by errors which must be taken into account when analysing the estimation procedures.

The possibility of comparing register data and high-quality survey data is of particular interest when establishing rules for treatment of total or partial non-response in registers, cf. subsection 6.2.

Supplementary sample surveys should also be used as a monitor, currently controlling the contents of register-based statistics. Otherwise, gradual changes in the concepts of the source registers may lead to false conclusions. Such changes may occur as a consequence of changes in law or in administrative practice.

The need for supplementary sample surveys has only been fully recognized in Denmark in recent years. Little experience has been gained as regards the practical use of such surveys.

#### 6.4. Longitudinal studies

A register-based statistical system with a unique identifier for each object creates new possibilities for statistical analysis where the objects are studied over a long period of time. This kind of analysis demands an identification system that is absolutely precise and stable. The scope for longitudinal studies is particularly bright in the person-statistics area.

The strong point of longitudinal analyses is the possibility of testing hypotheses about long run effects of the persons being exposed to some agent. The method is very well suited for research, e.g. in social medicine. Longitudinal studies may also yield valuable information for forecasting. Probabilities of transition from one situation to another (e.g. from education to employment) may be estimated and used for prediction of future development, e.g. future supply of labour.

The opportunities for longitudinal studies are far from being fully exploited. It is to be expected that work in this field will grow rapidly. However, many methodological problems will have to be explored. One question is how to organize statistical registers to make it possible to correlate (1) historical exposure data from a long period of time (e.g. a person's employment history during a lifetime), and (2) data on possible effects (e.g. cancer, death). Interpreting the results of such an analysis will also pose great difficulties.

#### 6.5. Documentation of statistical registers

Once a register-based statistical system is established, the potential for producing valuable socio-demographic statistics is enormous. Only a small fraction can be published, so other ways of dissemination must be employed. To derive full advantage of the potential, it is necessary to give consumers of statistics a clear understanding of what the system has to offer. This becomes increasingly important as the register-system is extended.

We need to find out how the documentation should be organized to reach this goal. Obviously, the documentation will be comprehensive and complex. It will probably be necessary to create a data-base system giving consumers direct access to the information on the structure and contents of the register-statistical system.

## SUMMARY

The purpose of the present paper is to identify and discuss some of the most important methodological problems encountered when building a statistical system based on administrative records.

In Denmark administrative records are the primary source of socio-demographic statistics. A coherent statistical system has gradually developed since 1970, and the 1981 Population and Housing Census is based exclusively on administrative registers. This has been possible because the administrative authorities keep a large number of edp-based registers with information on citizens. An essential precondition for the statistical utilization is the existence of a unique Person Number identifying each person. The number is used in practically all administrative systems, allowing automatic linking of data from different sources. Danish law for the protection of privacy recognizes the need for statistical utilization of administrative registers.

The development of register-based statistics poses a series of new methodological questions. So far, some of the problems have hardly been explored, and much methodology work is needed.

One question concerns the feasibility of extending the administrative records to include data for statistical use only. Such extensions are necessary to accomplish the goal of register-statistics, but it is difficult to secure the reliability of data that are not required for the purpose of the administration itself.

The problem of missing data is a serious one to register-based statistics. It is more complex than the non-response problem of traditional surveys and censuses, and some ideas for solution need to be analysed.

Supplementary data collection by traditional methods is required to monitor register-based statistics and to serve as a base for estimation procedures.

The systematic use of Person Numbers makes it possible to conduct longitudinal studies. Important results may be reached by this kind of analysis, but we are facing problems of organizing data and interpreting the results of longitudinal analysis.

Lastly, if consumers are to reap the benefits of register-based statistics, it is necessary to develop an efficient way to describe the contents of the total statistical system.

## ANNEX 1.

### Milestones in Danish register-based statistics since 1968

1. With effect from 1968, statistics of sales, purchases, etc., are compiled for most industrial classification groups on the basis of registered data from the VAT administration combined with data from Danmarks Statistik's business register.  
From 1971, similar statistics are introduced concerning employment.  
After these statistics have been established, decennial business censuses are no longer conducted.
2. With reference to 1st May 1970, register-based statistics are for the first time compiled for the total population in the individual municipalities, analysed by sex, age and marital status. In the period 1970-1973 the rest of the current population statistics are converted to CPR basis.
3. From 1970 the income statistics are converted to a register basis, following the introduction of the system of taxation at source.
4. In 1971 Danmarks Statistik establishes a "Mini Population Register" for use in developing register-statistical methodologies, including data matching, and for use in carrying out analyses and ad hoc surveys.
5. Following a decision made in 1971, Danmarks Statistik in 1973 takes over from the Ministry of Education the micro-level statistics of pupils and students.
6. In 1974 the Minister of Economic Affairs at the advice of Danmarks Statistik decides not to conduct a traditional population census for the mid-1970's.  
Instead a register-based population census (without housing data) is carried out for 1976.
7. With reference to 1st April 1977, building stock and dwelling stock statistics are for the first time produced on the basis of the recently established Central Register of Buildings and Dwellings. With reference to 1st January 1980, the first register-based housing census is conducted, combining data on dwellings and data on occupants.
8. With effect from 1979 it is decided at the initiative of Danmarks Statistik to extend the data contents of the tax authorities' registers by information on the permanent workplace of each employee.
9. On 1st January 1979 a continuously updated register concerning labour market affairs is established. The register (called CRAM) serves various purposes, including Danmarks Statistik's compilation of unemployment statistics.
10. With reference to 1st January 1981, the first full-scale population and housing census is conducted on the basis of registers.