

Regional Training on Producing Register-based Population
Statistics in Developing Countries
23 September – 31 October 2013

e-learning module: *Basic information and statistical background*
23 – 27 September 2013

Lesson 3

**Using administrative data in production of
population statistics; *register-based surveys***

References and complementary readings:

1. United Nations (2011): [Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices](#)
 2. United Nations (2007): [Register-based Statistics in the Nordic countries](#)
 3. Wallgren, A. and Wallgren, B. (2007), *Register-based Statistics: Administrative Data for Statistical Purposes*, John Wiley & Sons, Ltd, Chichester, UK.
-

Contents

1. What is a register-based survey?
 2. From administrative to statistical registers; register-statistical processing
 3. Pros and cons of registers-based surveys
-

1. What is a register-based survey?

Within national statistical offices (NSOs), three kinds of statistics are published; statistics based on sample surveys, statistics based on censuses and statistics based on administrative registers. It is most common to only differentiate between sample surveys and censuses, where the statistical office is responsible for the collection of the data. These two types of survey are dominated by the work to “*collect data*”.

This lesson deals with the third type of statistics which are based on administrative registers where, instead of collecting data through surveys and censuses, administrative registers from different sources are adapted (*data are procured*) and processed to be suitable for statistical purposes. This kind of survey is called “*register based surveys*”.

The starting point for any survey is a number of questions in connection to a specific area of interest. A survey is carried out to try and answer these questions. Simply described, the survey work consists of following phases:

1. Determining the research objectives and planning of the survey.
2. Procurement and processing of data.
3. Estimation, analysis of data and presentation of the results.

Phase 2 of a survey, the procurement of data, can be carried out in different ways:

- a) With own data collection using a *sample survey*.

Example: The Labor Force Survey is conducted in many countries. A new sample is taken monthly, with new data collection and reporting.

- b) With own data collection using a *census*.

Example: The traditional Population and Housing Census, in which all households are interviewed or asked to complete a questionnaire which is then processed by the NSO.

Because censuses result in the creation of a register, micro-data from censuses are also included in the system of statistical registers and can therefore form the basis for register-based surveys.

- c) Existing micro-data is used for a *register-based survey*.

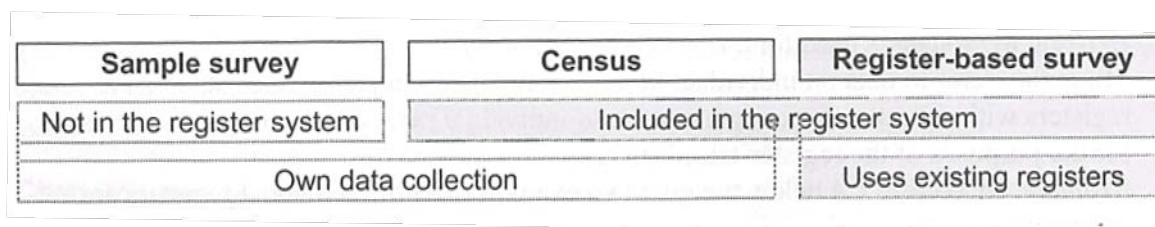
Micro-data refer to data on individual *units*. Existing administrative or statistical registers with data that, for example, refer to individual persons or enterprises are used for the purposes of the register-based survey.

Example: In Section 2 below we give example of Income and Taxation Register in Statistics Sweden.

Because these three types of surveys differ in terms of methodology, it is appropriate to differentiate them conceptually.

Data collection in a sample survey does not give rise to a register, as the microdata about the sample only consists of a small part of the surveyed population. Chart 1 compares the three types of survey that dominate at NSOs.

Chart 1. Comparison between the three types of surveys



2. From administrative to statistical registers; register-statistical processing

Administrative registers are created and delivered to a national statistical office

The original data formation is carried out in the authorities and organisations. The definitions of units and variables are adapted to administrative purposes. Every authority carries out controls, corrections and other processing that are suited to their administrative aims. When an authority delivers data to a NSO, further selections and processing may be carried out to meet the needs of the statistical office. The respective authorities also have metadata in the form of information on the definitions, data formation and quality. This type of information is also important for those receiving the data within the statistical office.

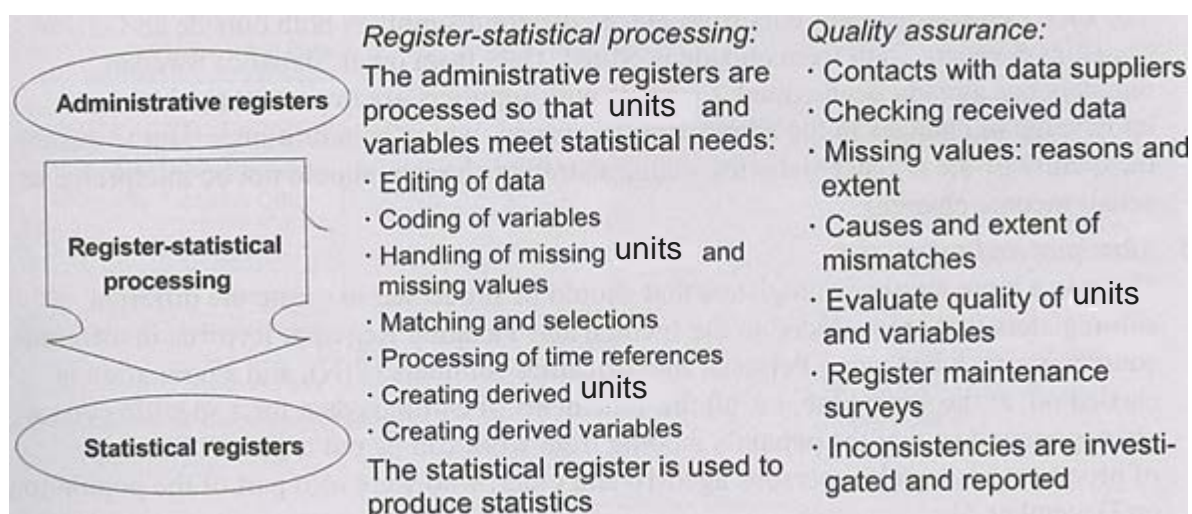
What happens when data is delivered to a statistical office?(the case of Statistics Sweden)

It is generally not a good idea to produce statistics directly from the received administrative registers because these are not adapted to statistical requirements. The unit sets, unit definitions and variables need to be edited and it will often be necessary to carry out some processing so that the register fulfills the statistical requirements for units and variables.

The register-statistical processing, which aims to transform one or several administrative registers into one statistical register, should be based on generally accepted register-statistical methodology.

The chart below shows the different elements included in statistical methodology work.

Chart 2. From administrative register to a statistical register



The next section describes how “Income and Taxation register” is created in Statistics Sweden. Though the example is from Statistics Sweden, but it illustrates general principles of creating a statistical register based on multiple administrative sources.

Statistics Sweden's Income and Taxation Register (I&T)

This register utilises many administrative sources. Many administrative variables are used to create important statistical variables. Besides these administrative sources it is necessary to use the register system at Statistics Sweden: the Population Register is used to define the population of the I&T Register, and important classification variables are imported from other registers in the system to the I&T Register.

1. Data formation at the National Tax Board

The annual income assessment is based on tax declarations from income earners and the taxation decisions of the local tax authority. Both the income earner and the tax authority use statements of earnings regarding salary, sickness benefit and interest that the employers, social insurance office and finance companies are responsible for. The National Tax Board ultimately compiles this information.

2. Microdata deliveries to the Income and Taxation Register

The Swedish National Tax Board annually creates databases that contain information on Sweden's population. The data files for one year - containing around nine million records, each with around 300 variables - are delivered directly to the I&T Register at Statistics Sweden.

3. Metadata to the Income and taxation Register

Record descriptions with variable names and variable definitions accompany the deliveries from the National Tax Board. Tax declaration forms, statement of earnings forms, taxation decisions, tax declaration instructions and instructions to employers are also needed to be able to interpret the data.

4. Editing of data

The I&T Register receives data from “eleven different suppliers” both outside and inside Statistics Sweden. Data from outside is edited. Data from other Statistics Sweden registers has already been edited. Contacts with suppliers are important to obtain knowledge of changes in the administrative system, which is in turn important to ensure the quality of the register statistics -administrative changes should not be interpreted as actual income changes.

5. Matching and selections

There is a large number of registers that should be processed to create the different sub-registers that are included in the I&T Register. Records in different sources are matched using Personal Identification Numbers (PIN), and aggregation is carried out at the same time so that the person's income from work can be put together. One type of processing is to select persons aged 16 and older, who were also part of the population on December 31

6. Derived units are created

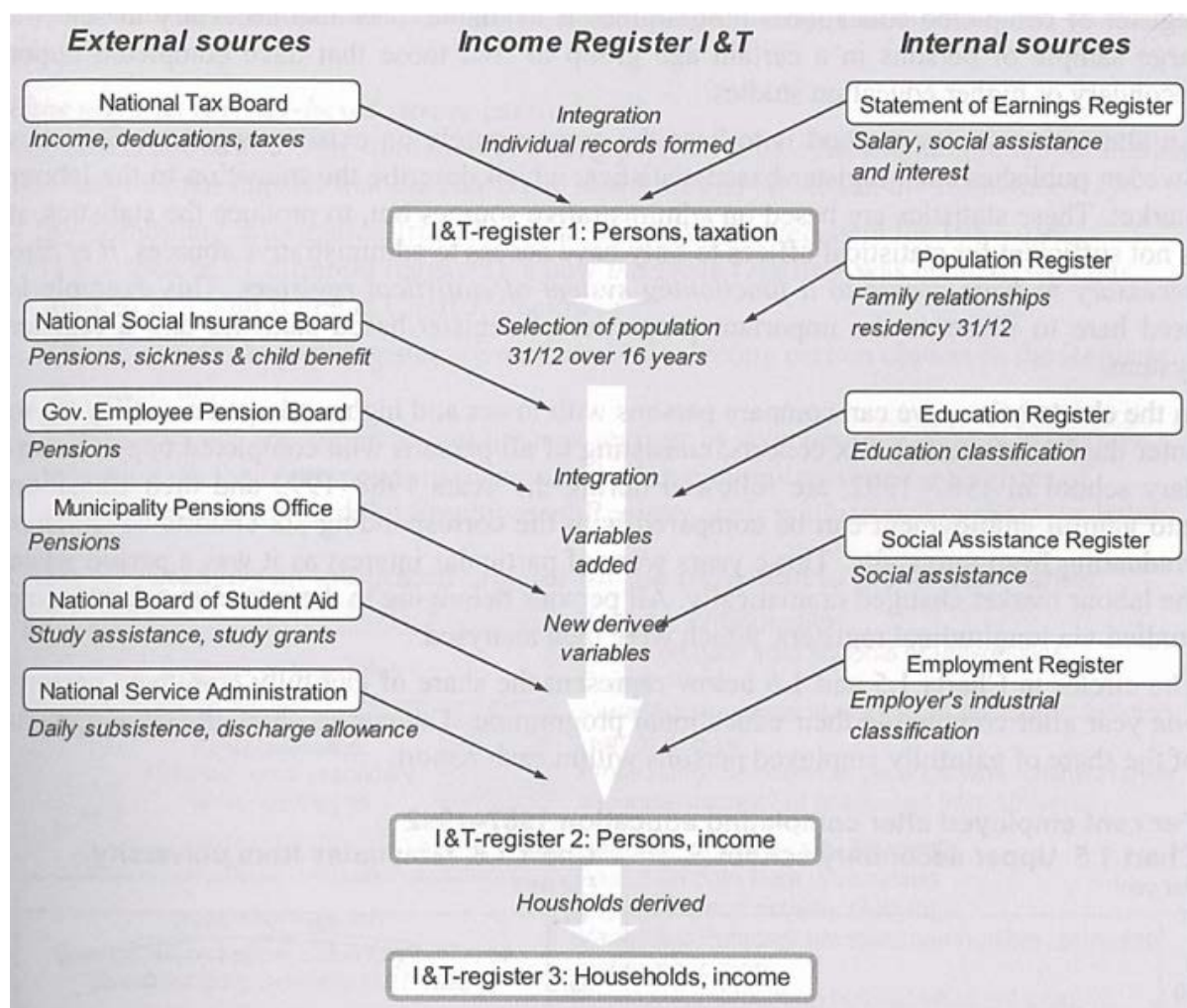
More information on certain relations helps to form household units. Between adults, the relations *married or cohabiting adults with children in common* result in that they are placed in the same household unit.

7. Derived variables are created

A large number of derived income variables are formed. For instance, the wage or salary amounts are aggregated from the different earnings data to become an individual's income from work. Every person's total income from work and capital plus transfer payments minus tax becomes the person's disposable income. For households, variables such as household type, number of consumption units and disposable income are formed.

The chart below shows how the I&T Register receives administrative data from a variety of different external sources and some Statistics Sweden registers.

Chart 3. Different data sources for the I& Register



3. Pros and cons of registers-based surveys

As mentioned, there is a common but often diffuse idea that statistics based on administrative data is of low quality. Is this idea justified for the administrative sources that a NSO uses? Normally a large part of register systems kept by NSOs is based on data from administrative sources. Would these statistics be of higher quality if those NSOs collect data themselves in parallel with the administrative registers? This is hardly the case – The NSO's own attempt to collect these data would be expensive, would increase the burden on the respondents and would likely produce data with more measurement errors.

For instance, in sample surveys and censuses, editing as a rule uses the collected data only. In register-based surveys, however, it is possible to compare variables from different sources, and this gives better possibilities to find and correct errors. Measurement errors and other kinds of inconsistencies can be detected if different sources are linked and compared.

Table below summarizes the advantages and disadvantages of the two survey methods, namely surveys based on data collection and register-based surveys which do not directly collect but procure data from different administrative sources.

	Advantages	Disadvantages
Surveys based on data collection: sample surveys and censuses	<p>Can choose which questions to ask</p> <p>Can be up-to-date</p>	<p>Some respondents ..</p> <p>... do not understand the question</p> <p>... have forgotten how it was</p> <p>... do not respond (nonresponse)</p> <p>... respond carelessly</p> <p>Burden on respondents can be high</p> <p>Expensive</p> <p>Low quality for estimates for small study domains (for sample surveys)</p>
Register-based surveys	<p>No further burden on the respondent for the statistics</p> <p>Low costs</p> <p>Almost complete coverage of population</p> <p>Complete coverage of time</p> <p>Respondents answer carefully to important administrative questions</p> <p>Good possibilities for reporting for small areas, regional statistics and longitudinal studies</p>	<p>Cannot ask questions</p> <p>Dependent on the administrative system's population, units and variable definitions</p> <p>The reporting of administrative data can be slow; the time between the reference period and when data are available for statistical purposes can be long</p> <p>Changes in the administrative systems make comparisons difficult</p> <p>Variables that are less important for administrative work can be of lower quality</p>