

# Toward an Integrated Statistical System Based on Registers





# Toward an Integrated Statistical System Based on Registers

Copyright © 2021 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/legal-code>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution and the use of the IDB's logo shall be subject to a separate written license agreement between the IDB and the user and are not authorized as part of this CC-IGO license.

Note that the link provided above includes additional terms and conditions of the license

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



Inter-American Development Bank  
1300 New York Avenue, N.W.  
Washington, D.C. 20577  
[www.iadb.org](http://www.iadb.org)

The Institutions for Development Sector was responsible for the production of this publication.

**External vendors:**

**Production Editor:** Sarah Schineller (A&S Information Partners, LLC)

**Editor:** Amy Scott (Nomad Enterprises, LLC)

**Design:** Gastón Cleiman

# Toward an Integrated Statistical System Based on Registers

Prepared for the Inter-American Development Bank by:  
Anders Wallgren and Britt Wallgren

## Abstract\*

This note describes how Latin American and Caribbean countries can join a revolution in statistical systems, moving from data collection based on geographic frames to one based on administrative registers, and the advantages of making this change. Northern European countries have already shifted from a traditional area frame-based statistical system to a register-based system, in which all surveys are based on statistical registers. Among the key advantages of the shift are: i) lower production costs; ii) potential for higher levels of geographic disaggregation and greater frequency; and iii) reduce the burden on informants by following the maxim of “ask once, use many times”. Evidence from Colombia, Ecuador, Mexico, and Peru points to the viability of this transition in the region. However, to take better advantage of the new strategy, countries should invest to improve the quality and coverage of their administrative systems and should create an integrated register system, allowing for efficient data use, and ensuring consistency and coherence across statistical registries.

**JEL Codes:** C1, C8, C19, H11

**Keywords:** statistics, statistical registers, administrative registers, surveys

\* Anders Wallgren and Britt Wallgren  
BA Statistiksystem AB, formerly of Statistics Sweden,  
E-mail: ba.statistik@telia.com

# Contents

<b>Preface</b>	<b>9</b>
<b>1 Basic Concepts</b>	<b>11</b>
1.1 What Is a Statistical Survey?	11
1.2 What Is an Administrative Register?	11
1.3 What Is a Register Survey?	12
<b>2 Why and When Should We Use Administrative Registers?</b>	<b>16</b>
2.1 Two Models of the National Statistical System	16
2.2 What Is an Integrated System?	17
2.3 Costs and Quality Comparisons	22
2.4 Combining Methods: Registers and Area Sampling	23
2.5 Why and When to Use Administrative Registers?	25
<b>3 How Should Administrative Registers Be Used?</b>	<b>26</b>
3.1 Use Identity Numbers and Protect Privacy	26
3.2 How Can Administrative Registers Be Used?	30
3.3 Uses of an Administrative Register	31
3.4 The Production Process Model and the Register System Model	33
<b>4 How Should We Make the Transition from Old to New Systems?</b>	<b>35</b>
4.1 The Production Process Model	35
4.2 The Register System Model	40
4.3 Quality Assurance	45
<b>5 Problems Related to the Population Register</b>	<b>48</b>
5.1 The Civil Register and National Statistical System	48
5.2 How to Find Duplicates	51
5.3 Undercoverage and Incorrect Residential Addresses	53
<b>6 Conclusions</b>	<b>61</b>
<b>References</b>	<b>63</b>



## Preface

Having timely and quality official statistics is fundamental for the design, implementation, and evaluation of a country's public policies. Traditionally, official statistics have been produced from three main sources of information: censuses, sample surveys, and administrative registers. Of these three sources, registers are used the least. In Latin America and the Caribbean (LAC), the national statistical offices (NSO) base the vast majority of their data production on censuses and surveys. However, such field work requires a high investment in terms of time and resources, as well as complex logistical planning.

Inspired by the successful experiences of countries such as Canada, Denmark, Spain, Norway, South Korea, and Sweden, which operate on the principle that the government should ask only once but use many times in order to reduce the burden not only on interviewers but also on interviewees, more and more of the region's NSOs are beginning to explore the enormous potential of administrative registers as a source of statistical data. This makes even more sense when considering that administrative systems continually generate data with a high degree of desegregation, in areas such as the civil register (births, deaths, and marriages), education, social security, health, employment, environment, migration, tourism, and so on.

Diverse factors have limited the statistical use of registers: the lack of coverage or standardization of the registers themselves, insufficient technical or technological capacity to process and integrate them, or legal frameworks that hamper NSOs' access to them. To address this need, for several years the Inter-American Development Bank (IDB) has been working closely with LAC countries and promoted various initiatives to improve the quality and coverage of administrative systems; strengthen NSO capacity to play a central role in the coordination, standardization, exchange, and interoperability of the registers; promote the creation of integrated and coordinated register systems; and set up modern statistical legal frameworks that allow for an adequate flow of data while also safeguarding its confidentiality. Some of the initiatives promoted by the IDB are: (i) the translation into Spanish and publication of *Register-based Statistics: Administrative Data for Statistical Purposes*, by Wallgren and Wallgren (2007), published as *Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos* (2012), in collaboration with Mexico's National Institute of Statistics and Geography (INEGI), which is also available in electronic version (Wallgren and Wallgren, 2016); (ii) the organization of training courses and technical consultancy provided by Wallgren and Wallgren, in Argentina, Belize, Barbados, Bolivia, Chile, Colombia, Costa Rica, Ecuador, El Salvador, and Mexico; (ii) the drafting of a

Generic Law on Official Statistics for Latin America, in collaboration with the United Nations Economic Commission for Latin America and the Caribbean (ECLAC, 2020), for the modernization of legal frameworks; and (iv) the provision of the online course Statistical Use of Administrative Records.

Thanks to these and to other efforts, several LAC countries have advanced in this area. For example, Brazil and Mexico have managed to link the census and household surveys with different administrative registers; Colombia and Costa Rica have included the identification number in the census to facilitate integration with other administrative sources of sociodemographic information; Chile, Ecuador, and Uruguay have carried out, or are planning to carry out, pilot census schemes based on administrative registers. To continue contributing to these advances, this publication provides a useful conceptual and methodological reference for those countries seeking to venture into or make further progress in designing, developing, and implementing an integrated register-based statistical production system.

**Karla Yee Amézaga**

Statistical capacity consultant  
Inter-American Development Bank

**José Antonio Mejía**

Modernization of the State Lead Specialist  
Inter-American Development Bank

# 1 Basic Concepts<sup>1</sup>

Two principles form the basis of our books on register-based statistics (Wallgren and Wallgren, 2012; 2014): the survey approach to administrative data and the systems approach. The survey approach discusses estimates, estimators, and quality as in books on sample surveys. The systems approach builds on the register system concept. Our books also discuss the production system at a national statistical office and the role of administrative registers in the design and development of that system. This document builds on that knowledge.

This first section discusses the following basic concepts: (i) What is a statistical survey?; (ii) What is a register?; and (iii) What is a register survey and what is the distinction between administrative and statistical registers?

## 1.1 What Is a Statistical Survey?

“Statistical Survey” is a central term used by statisticians at all national statistical offices. For many statisticians, however, the term “survey” is synonymous with the term “sample survey.” This will cause confusion when we discuss statistics based on administrative registers. To avoid this confusion, we follow the distinction between the different kinds of surveys used by Statistics Canada (2009). The term “survey” is used generically to cover any activity that collects or acquires statistical data. Those included according to Statistics Canada Quality Guidelines are:

- A census, which attempts to collect data from all members of a population.
- A sample survey, in which data are collected from a (usually random) sample of a population’s members.
- A registry survey [our term], in which data are derived from records originally kept for non-statistical purposes.

## 1.2 What Is an Administrative Register?

An administrative register is maintained to store records on all objects to be administered, and the administrative process requires that all objects can be identified. The following definition is valid for administrative and statistical registers:

---

<sup>1</sup> **Section 1** in this paper builds on **Section 1.5** in Wallgren and Wallgren (2014).

- A register aims to be a complete list of the objects in a specific group of objects or population. However, data on some objects can be missing due to quality deficiencies.
- Data on an object's identity should be available so that the register can be updated and expanded with new variable values for each object.
- Complete listing and known identities are thus the characteristics of a register.
- "Catalogue," "directory," "list," "register," and "registry" are different terms for the same concept. We will only use the term "register."

The following are examples of registers:

- Civic, civil, or national registration of the population in a country results in registers of citizens, births, and deaths.
- Income self-assessments from persons give registers of all taxpayers for a given year.
- In Sweden, enterprises with a turnover of SEK 40 million or more should report monthly. This gives monthly registers of all enterprises that have reported. For smaller enterprises, quarterly or yearly registers are obtained.
- All export and import transactions are registered by the customs agency. Monthly registers are created with all transactions for a specific month.
- A census file with data from a housing and population census is a register if the file contains identities of the persons in the file.

### 1.3 What Is a Register Survey?

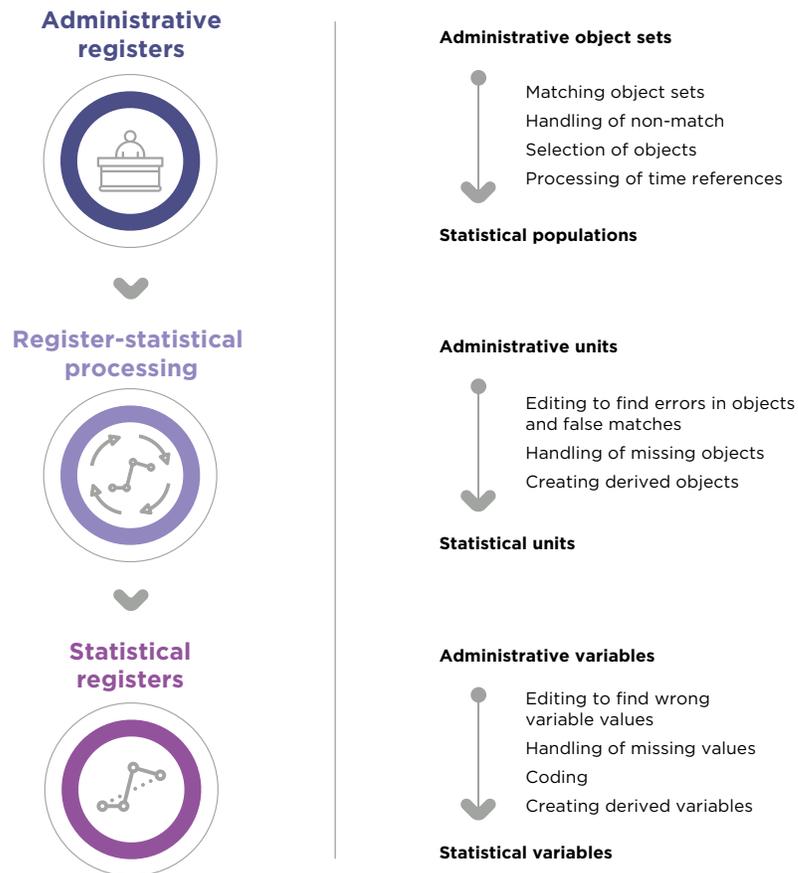
In **Section 1.1**, we explained that in a register survey, one or more statistical registers are created using administrative records or administrative registers. The original administrative data are generated in public administrative systems. Definitions of object sets, objects, or administrative units and variables are adapted to administrative purposes. Every authority carries out controls, corrections, and other processing suited to their administrative aims.

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 object sets, object definitions, and variables need editing. As a rule, it will be necessary to carry out a number of steps consisting of different kinds of register processing so that the register fulfills the statistical requirements for population, objects, or statistical units and variables.

**Figure 1** shows three important components of this work. We have found that people have a tendency to use administrative concepts as they are; in some cases this can be acceptable, while in other cases it is not. How population,

units, and variables of a statistical register are defined is important for the quality of the statistics that will be produced with the newly created statistical register. The register-statistical processing, which aims to transform one or several administrative registers into one statistical register, is described in **Figure 1** as a process where statistical populations, units, and variables are created.

**Figure 1.** From Administrative Registers to Statistical Registers



### Administrative Object Sets: Statistical Populations

A statistical Population Register can be created by combining five sources: the civil register for persons with identity numbers, the civil register for persons without identity numbers, the birth register, the death register, and the migration register. The statistical population at a specific point in time can

be created by using time references as dates of births, deaths, immigration, and emigration. Note that the administrative civil register and the statistical Population Register created by the statistical office would be different. The stock of persons will differ, and the variables in the two registers will differ. For example, if residential addresses in the statistical register are based on more sources than the addresses in the civil register, then the address variables will differ.

### Administrative Units: Statistical Units

In **Figure 2**, the administrative register with monthly transactions in a social insurance system is used to create three statistical registers for jobs, enterprises, and persons. All employers in the system register their transactions regarding all employees each month. In **Figure 2**, two kinds of identity numbers are used: *cédula* numbers that are personal identity numbers (PIN) and RUC numbers that are enterprise identity numbers (BIN).

In register 1 in **Figure 2**, the employer with RUC identity 111 has two employees, with *cédula* identities 11111111 and 22222222. Person 22222222 is also employed by the employer with RUC identity 222. All transactions with corrections and changes have been recorded in the system. When administrative register 1 has been delivered to the statistical office, statistical registers 2, 3a, and 3b are created. The administrative units “transactions” in register 1 are transformed into the statistical units “jobs” in register 2, “enterprises” in register 3a, and “persons” in register 3b.

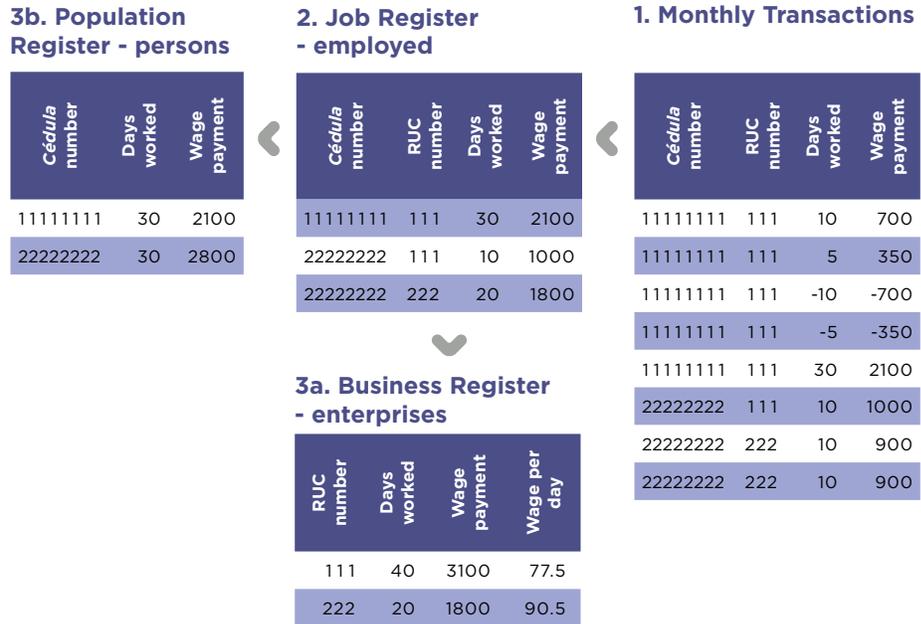
### Administrative Variables: Statistical Variables

There are two administrative variables in **Figure 2**: *days worked for transactions* and *wage payment for transactions* in administrative register 1 with monthly transactions. In statistical registers 2, 3a, and 3b, we have derived statistical variables “days worked” and “wage payment” for jobs, enterprises, and persons. Register 3a also has a derived statistical variable “wage per day” for enterprises.

In a sample survey, we define the statistical units, households, or establishments. The interviewers check to ensure that the intended kind of unit is interviewed. The questions and definitions in the questionnaire define the statistical variables. In a register survey, we must use the administrative units and the administrative variables. Sometimes we can aggregate administrative units and create statistical units, as in **Figure 2**. We can use the administrative

variables to create derived statistical variables, such as the variables in data matrices 2, 3a, and 3b in **Figure 2**.

**Figure 2.** From Administrative Units to Statistical Units



## 2 Why and When Should We Use Administrative Registers?

When a national statistical office starts to use administrative registers, it adds a new production factor to the production system. The opportunities provided by the old methods of conducting a census or a sample survey remain, but new opportunities will be added to the production system when administrative registers are now also used for statistical purposes. In this section, we analyze some of the consequences for the production system when administrative registers become available for statistics production.

### 2.1 Two Models of the National Statistical System

We compare here two different kinds of national production systems. The traditional system, where no administrative registers are used, is compared with the register-based system, where many statistical registers are developed using different administrative registers. Today, many countries are somewhere between these two extremes. Administrative registers are used to some extent, but the national production system is not yet completely register-based.

**Table 1.** Comparison between Traditional and Register-Based Production Systems

A traditional system: no registers	A register-based system: many registers
<p><b>Census</b> Based on an area frame. Interviewers knock on all doors and conduct face-to-face interviews.</p>	<p><b>Census</b> A number of statistical registers with yearly data replace the population census. Farm census: A register is used to create the frame; mail questionnaires are sent to all.</p>
<p><b>Sample survey</b> Based on an area frame. Interviewers knock on an area sample of doors and conduct face-to-face interviews.</p>	<p><b>Sample survey</b> A register is used to create the frame, mail questionnaires, telephone, or face-to-face interviews to a register-based sample (= "list frame" based sample).</p>
	<p><b>Register survey</b> With administrative registers, a statistical register is created.</p>

**Table 1** illustrates that the main methodological differences between the two production systems are the types of frames and data collection methods. Countries without a statistical population register can be classified as having a mostly traditional system, whereas countries that can conduct a completely register-based population and housing census can be classified as having a completely register-based production system.

As the Population Register is the most important statistical register, the classification into traditional or register-based systems can be based on the existence of a statistical Population Register. If the national statistical office has been able to create a Population Register, it will be relatively easy to create more registers.

## 2.2 What Is an Integrated System?

The Food and Agricultural Organization of the United Nations (FAO) has launched a global strategy to improve agricultural and rural statistics in developing countries. FAO (2010) discusses integrated national statistical systems. We have selected the following quotations from this report and have highlighted in italic text strongly related to what we mean by an integrated system.

### Box 1. FAO Global Strategy to Improve Agricultural and Rural Statistics in Developing Countries

The integration and underlying methodology considers the quality dimensions, which include relevance and *completeness*, accuracy, timeliness, accessibility, *coherence*, and *comparability*. The process of improving agricultural statistics will begin with the integration of agriculture into the national statistical system. This integration will be accomplished by the development of a master sample frame for agriculture to ensure relevance and completeness; its use *in implementing a coordinated data collection program to produce timely and accurate data that are coherent and comparable*. One of the shortcomings of current statistical systems in both industrialized and developing countries is that data are collected by sector, using different sampling frames and surveys. The division of data by sector leaves no opportunity to measure the impact of an action in one sector on another. In some cases, different organizations produce statistics for the same items, with different results, which confuse the data users. Surveys are often conducted on an ad-hoc basis with no links to a master sampling frame or the use of georeferenced units for data collection. It is therefore difficult to integrate data from various surveys for in-depth analysis with cross tabulation of variables. Integrated statistical systems can resolve many of these problems by avoiding duplications of effort, preventing the release of conflicting statistics, and ensuring the best use of resources. Concepts, definitions, and classifications become standardized, allowing more systematic data collection across sources.

**Source:** FAO (2010).

The parts of an integrated system harmoniously fit with each other. For a statistical system, this means that populations and variables in different surveys fit with each other. Important quality factors are completeness, coherence, and comparability. Ministries and institutes that conduct statistical surveys should also coordinate their data collection to produce statistics on society. FAO also wants microdata from different surveys to be combined or integrated at the micro level by record linkage in order to enable cross-tabulation of variables. According to FAO, there are many shortcomings with the present systems. Lack of cooperation between different actors leads to the publication of conflicting statistics and to duplication of efforts, which means that available economic and human resources are not used effectively.

Wallgren and Wallgren (2014) discusses the concept of a register system. Chapter 1 defines a number of principles for using administrative registers for statistical purposes. The following principles are related to the register system:

- System principles: All statistical registers should be included in a coordinated register system. This system will ensure that all data can be integrated and used effectively.
- Consistency principle: Consistency regarding populations and variables is necessary for the coherence of estimates from different register surveys.

In a national statistical system with a coordinated register system, it is possible to integrate all microdata in the system. In a coordinated system, the populations in the different registers are consistent regarding coverage, and variables in different registers do not have contradicting values. This is what is meant by an integrated statistical system. Of course, perfect consistency is difficult to obtain, but if coverage problems and variable inconsistencies are endurable, the system can be said to be an integrated system. Integration of microdata will become possible without disturbing problems regarding lack of consistency between populations and variables.

**Figure 3** illustrates a register-based production system. The basis is the register system consisting of four base registers (circles) defining the populations in the system and a number of statistical registers with important statistical variables. The lines in the figure illustrate the important links between units in different registers. We have also added two sample surveys: the Labor Force Survey and the Investment Survey.

**Figure 3.** A Register-Based Production System (Register System and Sample Surveys)

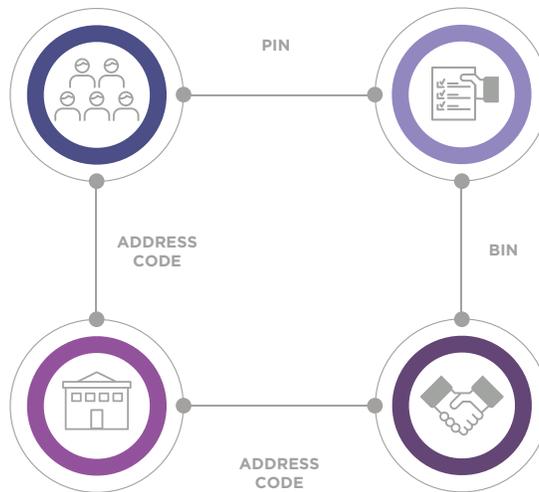
**Population Register**

**Sample surveys of persons  
Labor force survey**

Employment Register  
Education Register  
Income & Taxation Register  
Longitudinal Income Register  
Longitudinal Welfare Register  
Registers for medical research

**Activity Register**

Preliminary taxation system  
Social insurance system  
Compulsory school, pupils  
Upper secondary school, pupils  
Register of university students



**Real Estate Register**

Geographical database, GIS  
Land cover, Land use  
Register on buildings  
Register on dwellings

**Business Register**

Value Added Tax Register  
Annual Pay Register  
Yearly Income Tax returns  
Foreign Trade Register  
Farm Register

**Sample surveys of enterprises  
Investment survey**

**Notes:** PIN = identity numbers of persons; BIN = identity numbers of enterprises

If populations and variables are sufficiently consistent, the system in **Figure 3** is an integrated system. An integrated system has the following characteristics:

- Microdata from any register can be combined with microdata from all other registers. For example, employment, education, and income can be combined in a new register.
- Microdata from any sample survey can be combined with microdata from all registers. For example, the Labor Force Survey can use variables from

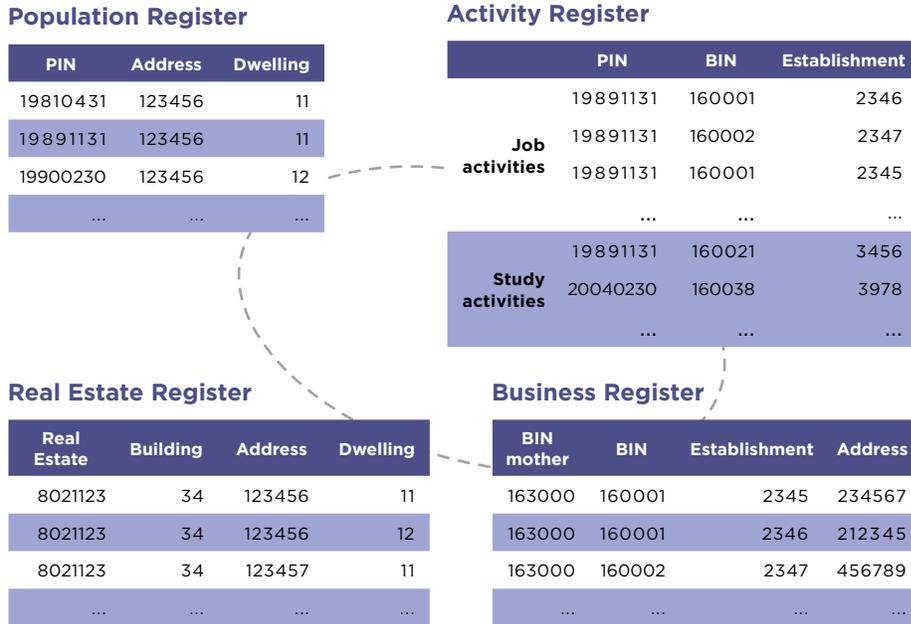
the Employment Register to improve estimation. It can also use variables from the Education Register instead of asking questions on education. This will improve quality and reduce costs.

In contrast to the integrability and flexibility of register data, microdata from a sample survey cannot be combined with microdata from another sample survey. This is because as a rule only a small number of units are included in more than one sample—there are almost no common units that can be combined. This explains why the systems approach is not needed in a traditional production system with no registers. It is sufficient to think of one sample survey at a time.

This fact explains why register-based statistics and sampling-based statistics require different approaches. The systems approach is fundamental for register-based statistics but is not at all required for sampling-based statistics. Books on sampling theory and sample survey methodology focus on discussions of *one* sample survey, not on a system of surveys. However, when we create statistical registers, we must think about many possible sources simultaneously, the many ways of using a source, and how to use different parts of the register system.

**Figure 4** shows statistical units and their identity numbers in the four base registers. With these identities, all registers can be linked within the system. As identity numbers of persons (PIN) or identity numbers of enterprises (BIN) are included in sample surveys, these can also be linked with the system. In the activity register the identity numbers, PIN and BIN, are combined. Activity is the statistical unit, and the same person can have many jobs and study activities during a specific period. The activity register is the link between the population and business registers. Thanks to this link, social statistics and economic statistics are parts of the same production system.

**Figure 4.** The Four Base Registers and their Links



A register-based system also opens new possibilities to organize work so that duplication of effort is eliminated. One team should be responsible for maintaining the Population Register and to create all populations of persons in the production system; another team should be responsible for all income variables in the system, and so forth. Such standardized populations and standardized variables are then used in all other surveys. Subject matter competence is deepened and more specialized, and the statistics produced will be more consistent and coherent.

New possibilities for quality assessment will also arise when the statistical production system becomes register based. Coverage problems will be found if populations in related registers are compared. The registers in **Figure 3** can be compared in the following ways:

- Are all addresses and dwellings in the Population Register and addresses in the Business Register also included in the Real Estate Register?
- Are all persons in the activity register also included in the Population Register, and are all employers in the activity register also included in the Business Register?

If similar variables in different registers are compared and if variables in a sample survey are compared with similar variables in registers, then errors

and inconsistencies in variable values will be found. One important example of this kind of work is to compare the coding of economic activity (ISIC or CIIU) in different sources. With the activity register, it will be possible to find the employers of the interviewed persons in the Labor Force Survey. The coding of economic activity in the Labor Force Survey can then be compared with the coding in the Business Register.

In summary, an integrated statistical system based on registers has the following advantages:

- It is possible to integrate microdata from different sources and create new combinations. In this way, existing data can be used more efficiently.
- Duplication of effort can be reduced and consistency improved.
- New ways of conducting quality assessment become available.

The creation of an integrated national system based on registers requires much work and new ways of thinking. This long-term project needs continuous support from managers and politicians. Cooperation with ministries and administrative authorities is also necessary for a successful implementation of the new methods for statistics production.

### 2.3 Costs and Quality Comparisons

Why should administrative registers be used for statistical purposes? There are two reasons for changing the national statistical system in this way: to reduce costs and to improve quality. When we work with survey design, we always want to achieve these two goals.

In **Figure 5**, two survey designs are compared. If we assume that **design 1** is a labor force survey in a traditional system and **design 2** is a labor force survey in a register-based system, then costs can be saved, as mail questionnaires or telephone interviews can be used in a register-based system instead of sending out interviewers to visit an area sample of households. In addition, quality can be improved, as sampling errors and nonresponse errors will be reduced when auxiliary variables in the register system are used to improve estimation.

**Figure 5.** Survey Design: Reduce Costs and/or Improve Quality

Three kinds of costs should be minimized. Financial resources can be saved, and the response burden will be reduced by the transition to a register-based system.

Traditional data collection will disturb the respondents in many cases. When people get a telephone call from the statistical office during the evening, this can be experienced as a loss of privacy. In an interview, the interviewer may have obtained confidential information that could be a threat to the privacy of the respondents. We can reduce these costs with register surveys.

In a register-based system, anonymous numbers to protect integrity should replace identities and access to microdata in the registers should be restricted. If these measures are not taken, costs as losses of integrity may increase when a register-based system is developed.

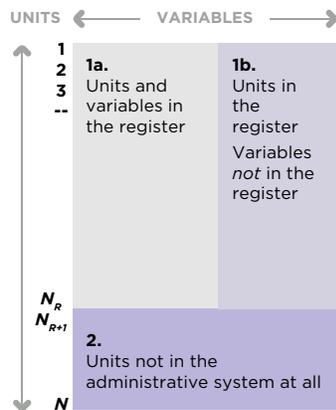
When discussing quality, we should remember that a register-based system opens new possibilities. Instead of an expensive traditional population and housing census, you will have a number of register surveys that will provide yearly statistics that did not exist before. For example, yearly statistics for small regions will only be possible in a register-based production system. It will also be possible to create longitudinal registers that will be very important for researchers. In these cases, we actually compare a situation with lack of important survey data in the traditional system with the quality of the new register surveys.

## 2.4 Combining Methods: Registers and Area Sampling

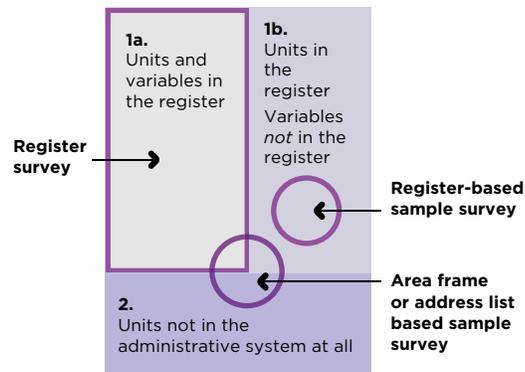
In a register-based system, we need both register surveys and sample surveys as two complementary methods. In Sweden, sample surveys are conducted when important variables cannot be found in administrative sources, and for short-term statistics when administrative registers are generated too late.

In many countries, the coverage of the administrative systems is not perfect. In developing countries, informal sectors can exist especially in rural regions where persons and farms are not included in administrative registers. Then it will be necessary to combine register-based surveys with area-based sampling. In addition, some administrative systems are not updated regularly. As a result, important variables in the registers, for example residential address, can be incorrect. The magnitude of such errors can be measured with area-based sample surveys.

**Figure 6.**  
Different Parts of the  
Target Population



**Figure 7.**  
Three Different Surveys



When a statistical register has coverage errors due to undercoverage, it can be combined with a sample survey based on an area frame. As the area sample theoretically has no coverage errors, the combined estimates can be unbiased. In a similar manner, a register-based sample survey that uses a register with undercoverage as a sampling frame can be combined with an area sample. In **Figure 6**, the target population consists of  $N$  units and the population is in two parts: *Part 1 (1a and 1b)* consists of the  $N_R$  units that are included in the register. *Part 2* consists of the  $N - N_R$  units that constitute the register's undercoverage.

The register survey in **Figure 7** can be combined with an area sample and the register-based sample survey can be combined with an area sample. In **Section 5.3**, we discuss how these area samples can be used for quality assessment and adjustment of coverage errors.

## **2.5 In Summary: Why and When to Use Administrative Registers**

Administrative systems in the public sector generate vast amounts of data that are stored in administrative registers. These registers are a resource that can be used for statistics production. This resource is growing in volume and, as the administrative systems gradually improve, the quality of the administrative data that are generated by these systems will improve. Using administrative registers for statistics production reduces costs and improves quality. Register surveys can replace censuses, and new kinds of regional and longitudinal statistics can be developed. The national system can be changed into an integrated statistical system with consistent and coherent statistical registers and sample surveys. During the transition period, it may be necessary to combine the register-based statistics with estimates from area samples.

## 3 How Should Administrative Registers Be Used?

Wallgren and Wallgren (2012; 2014) discuss the methods that should be used in the work with administrative registers. This section provides a short description of these methods and highlights methods that are new or different in a country that starts the work with registers. In such countries, the national statistical office previously had access to statistical tables (macrodata) produced by the ministries and institutes that are responsible for administrative registers. An integrated statistical system cannot be created with macrodata only. An integrated system should be based on coordinated statistical registers. The first important change will involve the national office gaining access to all microdata that is required to develop and maintain the national system of statistical registers. With macrodata only, it is possible to compare sources and find errors and inconsistencies, but with microdata, it is also possible to find the causes behind these errors and make corrections that reduce the errors.

### 3.1 Use Identity Numbers and Protect Privacy

In a traditional system based on area frames, the location of a statistical unit is the factor that determines whether a unit is sampled and interviewed. A sampled person's name or identity is not important for the production of statistics. Consequently, information regarding the identities of units is not used in the production process. Countries with register-based production systems use identity numbers in all surveys: in all of the censuses, sample surveys, and register surveys. The Population Register and the Business Register are used to create frames, and the statistical units are identified with unique numbers that are used in the production process. In sample surveys, when a sample of persons/establishments has been selected from the register, the identities are used to contact the persons/establishments by mail or telephone. In register surveys, microdata from different registers are combined to create the data set required for the specific survey. These combinations are created by record linkage using identity numbers as matching keys.

In a register-based system, three issues are important regarding identities:

1. The quality of the identity numbers. If this quality is deficient, there should be a long-term project to improve this part of the national registration system. In the short run, the identity numbers can be supplemented with names, birth dates, and birthplaces.

2. Record linkage will be an important part of the production process when the national statistical system becomes register based. The reason is that all of the variables needed for a specific survey are not included in only one source; rather, sources must be combined.
3. The production system must be designed to protect privacy by using anonymous numbers rather than real identity numbers and by guaranteeing restricted access to sensitive variables such as names and addresses.

### Quality of Identity Numbers

If an administrative register is used frequently for important administration, the quality of identities will be good. National identity numbers are used to make administration of taxation and welfare systems efficient and according to the legal rights of citizens. These administrative systems regularly undergo changes and improvements, and new systems are developed. When a country starts to use administrative data for statistical purposes, a new and important task for the statisticians is to take part in the work of developing the national administrative systems. The statisticians can contribute with general quality measures and provide information regarding statistical needs so that administrative data can also be used for statistics production. If the national registration system registers all newborns and all new immigrants, errors due to undercoverage and duplicates are prevented. Duplicates can be generated if persons register later in life—an earlier registration with a previous identity number may then have been forgotten.

### Record Linkage

In the Nordic countries, record linkage has been practiced on a large scale over many decades. Each country has a well-established system for personal identity numbers that is used in all administrative systems. The identity numbers have high quality and large amounts of administrative data exist where these identity numbers are used. Deterministic record linkage is used: If identity numbers agree, there is a link. If identity numbers do not agree, there is no link.

In a register-based system, record linkage is always done when a register is created. If, for example, four sources are combined to create one statistical register, then four record linkage operations are done. For some registers, more than 100 linkage operations are done; the important longitudinal registers that are used for research are examples of such registers. The conclusion of this is clear: the mismatch rate must be small in each record linkage operation; otherwise, the final register will be difficult to use.

In countries without good identity numbers, probabilistic record linkage is sometimes used to combine one sample survey and one register, or one census and one register. Combinations of matching keys are used: names, birth dates, and birthplaces can be used instead of identity numbers. Sometimes these variables have been recorded by hand on forms. Because the mismatch rates are often significant, this method is not suitable in a register-based statistical system. In a register-based system, populations and identities must be handled so that deterministic record linkage can be used.

At Statistics Sweden, the Population Register only contains persons registered as permanently residing in Sweden. All of them have valid identity numbers. This principle gives rise to some undercoverage, as there are persons belonging to the desired census population that are not formally registered. However, the benefit from this is that record linkage becomes very easy, as all records in the administrative registers with valid identity numbers can be used.

A country may change its identity number system or have parallel systems, and some persons may be assigned new numbers. The unit working with the statistical Population Register must handle this. A system is required that links old numbers to new numbers for persons that have changed identity numbers. In addition, different kinds of numbers should be transformed into the anonymous identity numbers that are used within the statistical office.

### Protecting Privacy

When a country changes its national system from a traditional to a register-based system, new methods aimed at protecting privacy must be developed. Identities can and should be handled in such a way that privacy is protected at least as well in the new register-based system as in the traditional system.

We first consider how administrative data looks and how this information is used in the administrative process. Assume that the tax form from *taxpayer X* shall be used to decide on taxable income and taxes to be paid. To check the information on the tax form, the clerks working at the tax authority compare it with the tax form for the spouse of *person X*. If these persons also own an enterprise, the tax form from the enterprise must also be checked, as there can be dubious transactions between *X*, the spouse, and the enterprise. Comparing with tax forms from the previous year is also a good idea.

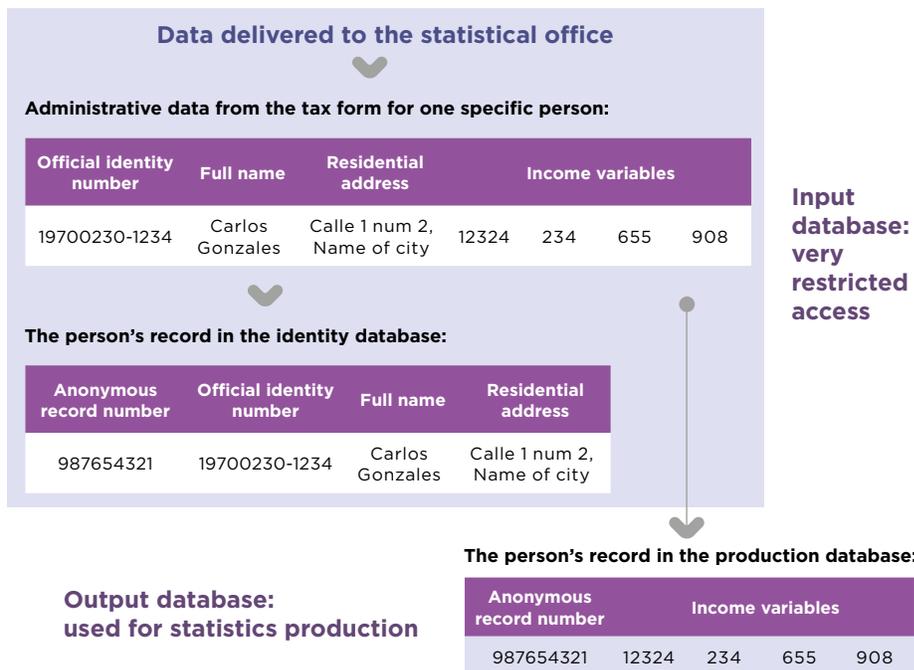
The situation described above is typical of the work with administrative data. The administrative IT system makes it easy to get access to all relevant

information for one specific person—in this case, six tax forms with perhaps 100 variables in each are combined for *taxpayer X*. Together with all of this data, names and addresses of persons and employers are given. With these data, the clerks at the tax authority decide: can we trust *taxpayer X*? If the answer is yes, taxable income and tax to be paid are determined. As there are many clerks working on checking tax forms, many people at the administrative offices have access to administrative data combined with official identity numbers, names, and addresses. These clerks easily recognize famous and well-known persons.

The work at the national office should be organized so that the use of official identity numbers, names, and addresses is minimized and highly restricted. In this way, privacy will be better protected than before. The work process is illustrated in **Figure 8**.

In a decentralized national system, statistics are produced at the ministries and institutes responsible for each administrative system. Clerks with access to microdata with easily recognizable names and addresses produce these statistics. When the national system becomes more centralized and register-based, this production is replaced with the work done at the national statistical office without access to names and addresses.

**Figure 8.** An IT System that Protects Integrity



One unit at the national statistical office, consisting of a small group of people, should be responsible for receiving all microdata that is stored in an input database with very restricted access. At this unit, names and addresses are removed and official identity numbers are replaced by anonymous numbers. After this work has been completed, the anonymized data are stored in the output database and the production database where those who need access to data for statistics production can use it for their purposes. Record linkage is carried out with the anonymous numbers in the production database.

The most sensitive information is the link between a person's identity and their address. Therefore, there should be a rule that the statistical registers in the output database only use information regarding the municipality or the relevant statistical region where a person lives.

### 3.2 How Can Administrative Registers Be Used?

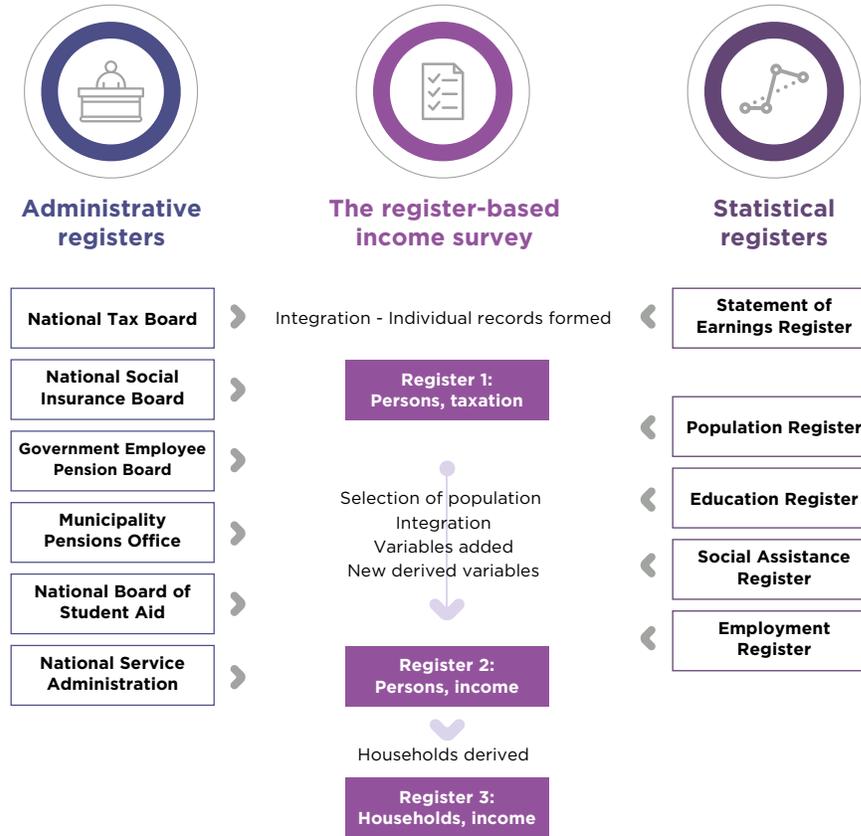
How are data from sample surveys used? After the data collection phase, there is one set of microdata. This set of data will be used in a specific way according to the survey's purpose. A well-established tradition of survey design builds on these premises. There is a risk that we think in a similar way when we want to use administrative registers for statistics production—we gain access to one administrative register and try to discover one way to use this set of data. With this traditional way of thinking, the full potential of the new sources with administrative data will not be realized. The design of register surveys should instead be based on the following principles.

We define a register survey as the work that is required to create one or more statistical registers for a specific purpose. The following facts explain why register survey design is more complex than sample survey design:

- One administrative register can be used for many register surveys and statistical registers.
- For one register survey, many administrative and statistical registers may be needed.
- The statistical register may be used by many others for other surveys.

The Swedish register-based income survey in **Figure 9** illustrates this point. A number of external administrative registers are used as input. In addition, a number of internal statistical registers from Statistics Sweden's register system are used. Three statistical registers are created with the sources and constitute the microdata that are used for the income survey. Some of the sources used as input to the income survey are also used for other register surveys. Tax forms from the National Tax Board and statements of earnings are also used for enterprise statistics.

**Figure 9.** The Swedish Register-Based Income Survey



### 3.3 Uses of an Administrative Register

When the statistical office gains access to new sources, how should it use these new administrative registers? How should the existing surveys be modified? To answer these important questions, the administrative sources should be analyzed by experienced subject-matter specialists and methodologists with a good overview of the statistical production system.

An administrative register or source can be used in four different ways. The first way, using the administrative register alone, will not contribute to an integrated system, but the other three ways of using an administrative register will improve the production system so that it becomes more integrated.

### 1. Completely Alone

If the source has good coverage and the variables in the source are of good quality, then the source can be used alone for producing statistics. Example: In a birth registration system (birth), the identity of the mother (identity number, name, address, and birth date) and medical data for the baby and the birth are registered by the hospital staff. Here, we have only data regarding one part of the female population—the new mothers. In addition to medical statistics, this register can be used to produce statistics on number of births by region and age of the mother.

### 2. Alone, But Combined with a Base Register

The Population Register and the Business Register are two important base registers used for all surveys regarding persons or enterprises in a register-based production system. If an administrative register or source is combined with a base register, the quality can be checked and improved. It will then be possible to produce consistent and coherent statistics. The base register contains important classification variables that can be combined with the administrative register. Example: If the birth register and the Population Register can be linked with the identity of the mothers, this combined register can be used to compare new mothers with all other women that are not new mothers and produce statistics on number of births per 1,000 women by region and age of the mother.

### 3. Combined with a Base Register and Other Administrative or Statistical Registers

In many cases, an administrative register does not have sufficient coverage and/or the variable content is too limited. Then the source cannot be used alone for statistics production. However, if many sources are combined, it may be possible to use the combined data set for register-based statistics. Example: If the birth register, the Population Register, and the statistical education register can be linked with the identity of the mothers, this combined register could be used to produce statistics on number of births per 1,000 women by region, age, and level of education of the mother.

#### 4. To Improve Other Surveys (i.e., to Improve the Production System)

If registers are compared, differences regarding coverage may be detected. Once these errors have been found and understood, it will also be possible to correct or reduce the errors. Wallgren and Wallgren (2015a) discusses such coverage errors as well as how to find and reduce errors and inconsistencies between classification variables that were found when we compared economic activity and sector in the Swedish Labor Force Survey and the Structural Enterprise Statistics survey. Example: The birth register noted above can be used to improve the quality of the statistical Population Register in three ways:

1. Address information regarding new mothers in the birth register can be used to update addresses in the Population Register.
2. Some mothers may not be registered in the administrative Population Register. This undercoverage can be reduced if data from the birth register for these mothers are added to the statistical Population Register.
3. Newborn babies can be included in the statistical Population Register.

All register surveys in countries with register-based statistical systems (e.g., the Nordic countries) use a base register as in methods 2 and 3 above. New register countries that have not yet developed good base registers will start with register surveys of the simple kind, noted in method 1 above. When base registers have been developed, it will be possible to create register surveys according to methods 2 and 3. Up to now, method 4 has not been practiced even in the Nordic countries; a common pattern is that everyone is fully occupied with their own survey and has no time to make comparisons with other surveys.

### 3.4 The Production Process Model and the Register System Model

The Generic Statistical Business Process Model (GSBPM) has been discussed at some national statistical offices. The focus of this model is on the different parts of the production process for one survey. **Figure 10** gives an overview of the process and each of the seven parts consists of many sub-processes. The main idea is to standardize methods by checklists, instructions, and common IT tools.

**Figure 10.** Parts of the Survey Production Process



The process model is applicable for sample surveys and censuses but does not support the systems approach that is essential for register surveys. The important work with creating, maintaining, and using the register system is not included in the process model. The model illustrates that the systems approach is not needed in a traditional production system with no registers. It is sufficient to think of one sample survey at a time.

In contrast to the philosophy behind the GSBPM, the register system model aims at standardizing populations and variables so that surveys become consistent and coherent. Data can be used efficiently in an integrated register system, which reduces costs, and populations and variables can be compared, which gives opportunities to find and reduce errors. **Figures 11 and 12** in **Section 4.1** illustrate process models that are suitable for register surveys.

## 4 How Should We Make the Transition from Old to New Systems?

A traditional statistical system based on area sampling can be more or less integrated. FAO (2010) points out that if all agricultural surveys use the same master sampling frame, and if the definitions of variables follow common standards, then the system is more integrated. However, it will never be possible to integrate microdata from different sample surveys even if these surveys are consistent and coherent.

When data from administrative registers are added to the system, new problems arise, but also new possibilities. The register data will differ from the sample survey data, which means that inconsistencies are now apparent. In addition, different registers will have different coverage and conflicting variable values and similar variables with different definitions. However, it is now possible to combine and compare microdata from different sources. Errors and inconsistencies will be found, and corrections can then be made. In this way, an integrated register system with consistent populations and variables can be developed gradually. Nevertheless, this will require much work, and many difficult problems must be solved.

In terms of organizing work when you have access to administrative registers, we use two different models when we discuss this. The production process model<sup>2</sup> in **Figure 10** will be used to describe how the work within the statistical office should be organized. The register system model (see **Figure 3**) will be used to describe the use of different administrative registers in developing an integrated system of statistical registers. Methodological issues related to the production process are discussed more thoroughly in **Section 5**.

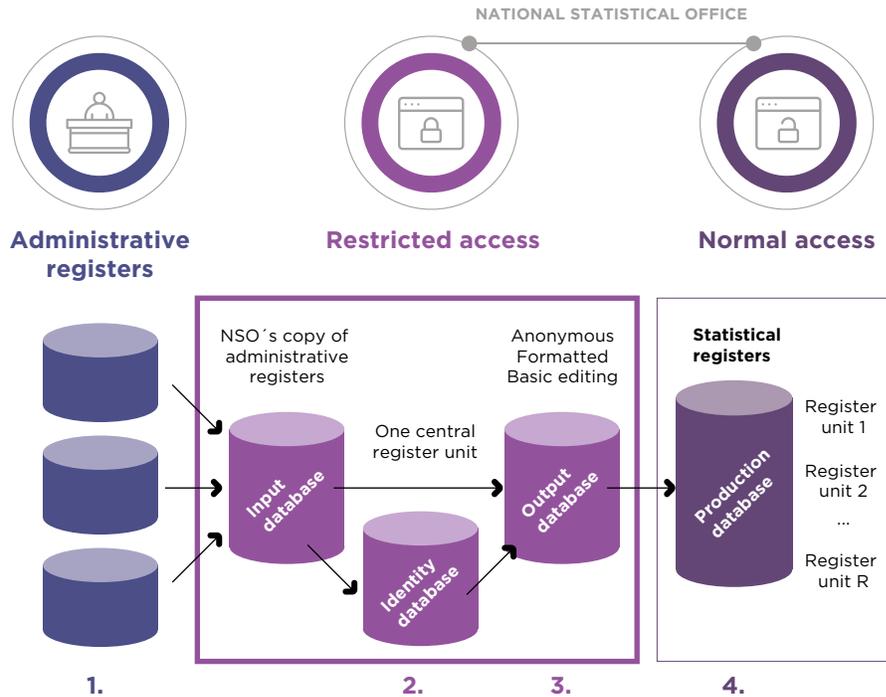
### 4.1 The Production Process Model

The production process from administrative into statistical registers is illustrated in **Figure 1**, where statistical methods are highlighted. The same process is also illustrated in **Figure 11**, which shows the organization of the work. We structure the process in four steps numbered 1-4 at the bottom of the figure.

---

<sup>2</sup> This model was developed together with David Chaín, INEC, during our visit to Ecuador in 2015.

**Figure 11.** The Production Process: Transforming Administrative Registers into Statistical Registers



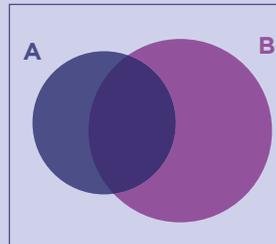
1. Administrative registers are delivered by administrative authorities to the statistical office. The selection of units and variables to be submitted must correspond with the statistical requirements. Cooperation between the administrative authorities and the statistical office should be well established and permanent. For updating the Population Register and the Business Register, only new and changed records should be submitted.
2. With the identity database identifying variables such as personal identity numbers, other kinds of identity numbers and names are replaced by anonymous record numbers. Only these record numbers are transmitted to the output database. Records with different personal identity numbers for the same person are linked with the same anonymous record number.
3. The variables in the input database are transformed and formatted so that all variables in the output database comply with the standards and formats used at the statistical office. In addition, basic editing of the variables in each administrative register is done in the output database.
4. The main work of creating statistical registers is done in the production database. Different sources are combined by record linkage, register populations are created, derived units and derived variables are created, and consistency editing is carried out.

## Contacts with Administrative Authorities

For the statistical office, the contacts with administrative authorities are necessary for many reasons. The statistical office needs not only microdata but also metadata with definitions of all administrative variables. The administrative process that generates the administrative registers must be understood—how has the administrative object set been generated? These object sets will be used to create statistical populations. Therefore, the statisticians must understand the administrative system to be able to judge the relevance of the data.

Contacts with the statistical office are also important for the administrative authorities. The authorities will gain new quality information. New kinds of comparisons with other sources can be made with the statistical register system. Lack of statistical quality will also indicate lack of administrative quality. Undercoverage is not only a statistical problem, it is also an administrative problem. Duplicates or incorrect residential addresses are also administrative problems.

The authority often has much more data than the statisticians expect. The rectangle to the right represents the data that the administrative authority has. Area A represents the data that the authority thinks that the statisticians need, and the dotted area B represents the data that the statistical office should get. To avoid misunderstandings, the selection of data to be submitted must be decided after a thorough discussion between the two partners.



Working groups from the statistical office should handle cooperation with administrative authorities. These working groups should consist of staff with competence in the subject matter, on register-statistical methods, and on IT. Cooperation between statisticians and different ministries and administrative authorities should be a long-range project for the purpose of developing the national administrative and statistical systems. **Section 5.1** discusses the civil register and statistical system.

## The Central Register Unit

The identity database is developed for two purposes: to protect integrity and to facilitate register processing. Integrity will be protected as names and personal identity numbers will be removed. The staff working with register-based

statistics will only have access to anonymous record numbers when they create their statistical registers.

However, there can be parallel systems of identity numbers. The identity database will then make the production process easier and the risk of mistakes will be reduced. This is because all data used in the production database have been edited so that complicated identifying variables such as names and different kinds of identity numbers have been replaced by a unified system of record numbers that are easy to use.

A central unit with a limited number of staff should be organized and given responsibility for receiving all administrative registers. Only this unit should have access to the servers and databases where copies of the administrative registers are stored.

The central register unit is also responsible for replacing names and official identity numbers with anonymous record numbers. Only the central unit, therefore, has access to the identity database where the links between identities and anonymous record numbers are stored.

The central unit is also responsible for handling identities used in register-based sample surveys. A register-based sample survey of persons should be conducted as follows:

- A random sample of persons is drawn from a frame created with the Population Register. This creates a list with anonymous record numbers for the sample.
- This list is sent to the central unit. It replaces record numbers with names and addresses that are used to send out questionnaires or gives names and addresses to interviewers. The combinations between names/real identities and record numbers are the most confidential part of the system and should never exist outside the identity database.
- The collected data is delivered to the central unit, which removes names and addresses and replaces them with the anonymous record numbers.
- The anonymous data set is delivered to the unit responsible for the sample survey.

Apart from handling identities, the central register unit is also responsible for two important tasks in the processing of administrative registers:

- Transforming all administrative data into the formats and standards that should be used at the statistical office. If, for example, three different administrative sources in the input database define sex in different ways—M/F, 0/1, or 1/2 for men/women—then in the output database, all these registers should use the same codes (e.g., 1 for men and 2 for women). This

kind of standardization will become much more important when the production system is changed from being area frame based to register based. In the register-based system, we can combine microdata from different registers and surveys. Then, the same formats and standards should be used in all registers and surveys.

- Basic editing of each administrative register. This editing consists of two tasks: searching for duplicates and finding and correcting erroneous variable values. This is described in Wallgren and Wallgren (2014: Section 9.1.1). Handling duplicates is also discussed in **Section 5.2** below.

### Do It Right from the Beginning!

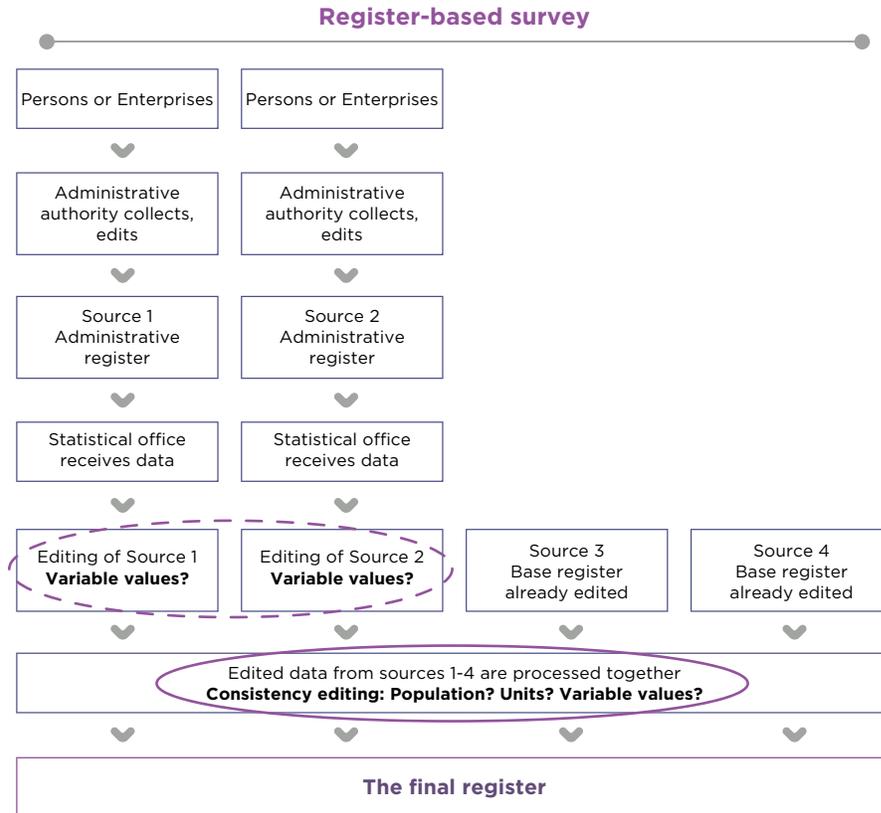
The identity database and the standards and formats to be used for the variables in the system should be established early in the process of modernizing the statistical system. If this rule is not followed, expensive IT systems will be developed that use original identities and formats in the administrative sources. When people later understand that privacy must be better protected and that standardization is necessary, changing these IT systems will be quite costly.

### The Production Database

Different register units in the production database create the statistical registers. Each unit can import the administrative registers it needs from the output database. It can also use the statistical registers in the system—all statistical registers in the system are stored in the production database. The only identity of persons is the anonymous record number used for record linkage. The responsible register units create statistical populations, units, and variables. These populations, units, and variables are then checked by consistency editing, which results in the discovery and reduction of errors and inconsistencies.

As shown in **Figure 12**, there are two kinds of editing in the creation of statistical registers. The first (dotted ellipse) is done in the output database by the central register unit when it edits each administrative register in the output database. After searching for duplicates, the editing is done with the same methods that are used for editing of sample surveys. The aim is to find and correct erroneous variable values.

The second kind of editing (solid ellipse) is the consistency editing that is done in the production database by the units responsible for each statistical register. Many different kinds of errors and inconsistencies will be found when the different sources are compared. Not only erroneous variable values will be found, but also coverage errors and errors related to statistical units.

**Figure 12.** Two Kinds of Editing in the Creation of Statistical Registers

The editing work that is done in the output database and the production database is the main part of the work with quality assurance of register-based statistics. Documentation of the results of the editing work is important. This will make it possible to monitor changes in the quality of administrative data over time. Quality assurance is discussed in **Section 4.3**.

## 4.2 The Register System Model

Statistical registers are created in the production database. The register system model is used to discuss the following topics:

- How to organize the work with respect to subject matter competence.
- How to organize the work to obtain consistency.
- How to do record linkage.
- How to use the system for survey design.

### Subject Matter Competence

As shown earlier in **Figure 3**, the register system model is structured by type of statistical unit: persons, activities, enterprises, and real estate or land. This structure is related to different subject matter fields. For example, economic statistics are related to the Business Register. However, the relation is rather weak, as economic statistics require data from all four parts of the system. Moreover, statistics on persons require data from the whole system.

How to organize the work with respect to subject matter competence? Let us take the four registers with data on persons as an example: the population, employment, education, and income and taxation registers. Annual registers are created every year: the population of persons as of December 31, and employment status, education, and income for the population of persons as of December 31. These data can be structured in four different registers or in one comprehensive social statistics register.

**Table 2.** Four Registers or One Comprehensive Register?

Population register	Employment register	Education register	Income and taxation register	Social statistics register
PIN	PIN	PIN	PIN	PIN
Age	Age	Age	Age	Age
Sex	Sex	Sex	Sex	Sex
Municipality of home	Municipality of home	Municipality of home	Municipality of home	Municipality of home
	Employment status	Highest level of education	Income from work	Employment status
	ISIC or CIU	Type of education	Income from ...	ISIC or CIU
	Municipality of work		Disposable income	Municipality of work
				Highest level of education
				Type of education
				Income from work
				Income from ...
				Disposable income

There are two questions here: how should the work be organized and how should the data be stored? Our answer to the first question is that the work should be organized so that the role of subject matter competence

is strengthened. In the example, there should be four register units with well-defined responsibilities:

- The population survey unit is responsible for population statistics and the Population Register.
- The employment survey unit is responsible for labor market statistics, the employment register, and the social insurance register.
- The education survey unit is responsible for education statistics and the education, compulsory school, upper secondary school, and university registers.
- The income survey unit is responsible for income statistics, the income and taxation register, and the longitudinal income register.

Each of these register units requires different subject matter competence; data are delivered by different administrative authorities and have different kinds of users. It is therefore preferable to organize four different register units when statistics production has become established.

Our answer to the second question is that the data can be stored in four independent registers, with some redundancy, or it can be stored in one comprehensive register with no redundancy. Based on our experiences from Statistics Sweden, we recommend the simplest organization with four independent registers where each register unit is responsible for its own database. The redundancy will be insignificant because the employment, education, and income registers will only import a small number of variables from the Population Register.

This section stresses the importance of subject matter competence. When the statistical system becomes register based, the role of this kind of competence changes. Contacts with administrative authorities and understanding of national administrative systems are new and important areas where this competence is required. How should the new statistical registers be designed to give us important information about society? Subject matter competence is needed.

The register system is structured by populations and variables—specifically, by the content of the registers. This means that different parts of the register system are linked to different kinds of subject matter competence. The work with statistical registers should be organized to promote this subject matter competence.

During the 1990s, many important new statistical registers were created at Statistics Sweden. Subject matter specialists led the work. The register units were free to contact authorities and future users of the new statistics and use

their creativity to develop the new registers. This was a decentralized organization where entrepreneurs had the freedom to create new statistics.

### Consistent Registers

In a perfectly integrated production system, different registers and surveys are consistent regarding coverage and variables. The register system model can be used to illustrate the conditions that are required to obtain consistency. The work must be organized to achieve this consistency. Managers must be involved. When errors and inconsistencies are found, decisions must be made regarding which register should be corrected or changed and how the costs for the improvements should be allocated.

Register-based statistics require teamwork between different units regarding populations and variables. If one unit works with monthly statistics and another unit with yearly statistics, this coordination may be forgotten due to the difference in reference periods. Supervising and strengthening this teamwork will be an important task for managers. We give two examples of this kind of cooperation: four yearly registers and an employment register.

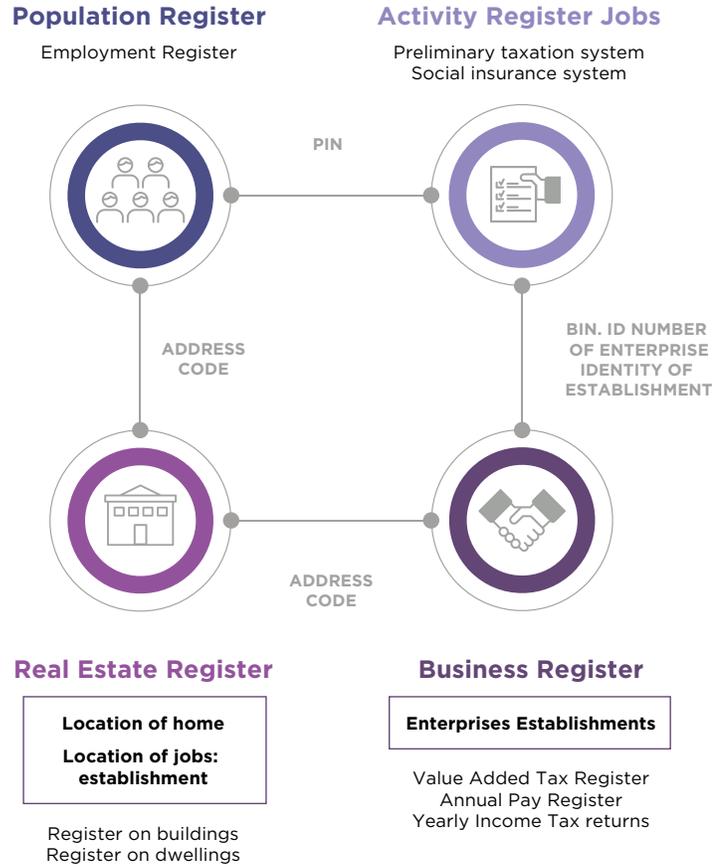
**Four Yearly Registers.** First, the unit responsible for the Population Register creates the population on December 31. Thereafter, the units responsible for the other registers take this register population with basic classifications by sex, age, region of residence, and others and add their own variables, such as employment, education, and income (see **Figure 13**). In this simple way, we obtain perfect coordination regarding populations and variables.

**Figure 13.** Four Yearly Registers



**Employment Register.** This register requires cooperation with politicians and cooperation between different departments at the national statistical office as shown in **Figure 14**.

**Figure 14.** Employment Register



First, employers with more than one establishment must report each employee's place of work. A political decision may be necessary to make this reporting a part of the social insurance or taxation system.

Second, data from all parts of the system are needed to create the employment register. Five identities will be used for record linkage: identity numbers for persons, enterprises, and establishments, and address codes for the location of residences and establishments. The unit responsible for the Business Register must create a register of all establishments. The main purpose of this register is to support the employment register. Cooperation is necessary to obtain consistency regarding populations—the population of employees and self-employed should be consistent with the Population Register and the population of employers and self-employed should be consistent with the Business Register.

## Record Linkage and Survey Design

When record linkage is discussed in scientific publications, it usually refers to record linkage of one set of data for persons with another set of data for persons. The register system model shows that in a register-based system, record linkage is more advanced and provides more possibilities. In such a system, persons can be linked with either enterprises or dwellings, and activities with either persons or enterprises.

This characteristic of an integrated system creates new opportunities for designing surveys. The employment register described in the previous section is an example where data from many parts of the register system are combined. Another example is the longitudinal register for research that Statistics Denmark (Stender, Thorsen, and Andersen, 2015) has created. A large volume of data on job and study activities and participation in different welfare programs is used to describe the population's attachment to the labor market.

Subject matter competence is required to create registers of this kind, and it is important to work together with academic researchers. This subject matter analysis defines the desired content. Subsequently, advanced register processing of many sources is required to create the desired register.

### 4.3 Quality Assurance

The statistical Population Register is the most important part of a register-based production system. Without this base register, it is difficult to create important register-based statistics; with this register, it will be easier to continue and develop other parts of the system. Therefore, a project to develop a statistical Population Register should be started early when a country has decided to develop register-based statistics. When the first version of this register has been created, the next step is to evaluate the register. Existing area-based sample surveys could be used for this purpose.

Administrative population registers in the Latin American countries we visited have two main flaws: undercoverage of persons that have not registered and outdated residential addresses. We discussed this in **Section 2.4** and we continue the discussion in **Section 5.3** and describe how area sampling can be used to estimate these errors.

Another kind of quality assurance is the analysis of the usability of each administrative register. Wallgren and Wallgren (2014: 129, 176, 273-277) describes four groups of quality indicators that can be used for this purpose.

The first group of indicators measures relevance. They are used to analyze metadata regarding each administrative register. This information consists of the documentation that has been created by the administrative authority responsible for the source. Administrative forms, supporting instructions, and interviews with the data suppliers should be used for this analysis. The relevance of the administrative population, units, and variables are assessed based on these metadata.

The second group of indicators measures accuracy. They are based on the analysis of microdata from each administrative register. The quality of matching keys, the existence of duplicates, missing values, and incorrect values are documented.

The third and fourth groups of indicators are based on comparisons with other sources and registers in the system. Coverage issues and inconsistencies between variables are analyzed with these indicators. Laitila, Wallgren, and Wallgren (2012) introduce these four kinds of indicators. A case study shows how they can be used to evaluate the quality of an administrative register.

## 5 Problems Related to the Population Register

The quality of the statistical Population Register depends on the quality of the administrative registers used as input. Administrative registers with poor quality can be compensated to some extent by suitable statistical methods where information from more registers is combined.

Wallgren and Wallgren (2014) discusses the variables that should be included in the statistical population register. How this register is updated with new information is also illustrated. Identifying variables, reference variables with links to other kinds of statistical units, and time references<sup>3</sup> with dates for demographic events are important variables in the population register.

In this section, we first discuss the national registration system, which determines how persons are registered in the civil register.<sup>4</sup> How can the registration system be improved? Creating an efficient register-based statistical system may require improving the routines used for creating and maintaining the civil register.

The second issue we discuss is how to find duplicates and how they should be treated. A person can have records from different points in time, and this information should be kept in the Population Register system so that changes over time can be tracked.

Finally, we discuss undercoverage in the Population Register and the problem with residential addresses that have not been updated. An area-based sample survey should be conducted regularly to monitor these quality problems. It may be possible to adjust for these kinds of errors with methods similar to the methods that are common practice for correcting estimates for nonresponse.

---

3 We recommend that time references have the format YYYYMMDD. Time variables defined in this way are ordinal scale variables and can easily be used for defining different register populations.

4 In all countries, one authority is responsible for the administrative population register. Here, we call the register the “civil register” (registro civil) and the authority the “registrar’s office.” We use the term “cédula number” for the personal identity number (PIN).

## 5.1 The Civil Register and National Statistical System

The registrar's office is responsible for registration of persons and for the system of identity numbers used in the country. As the population of persons is the most important target population, the statistical population register will be the most important register in the modernized national statistical system. When social statistics become register based, it may be necessary to improve the routines used for the registers handled by the registrar's office. Otherwise, it will not be possible to create register-based statistics of sufficient quality. The registrar's office and the national statistical office must work together in a long-term project for this purpose.

In the Nordic countries, the administrative population registers and the systems for identity numbers were good from the beginning. When the Nordic countries started to produce register-based statistics during the 1960s, they were able to use just one source for the population register, as the quality of that source was good. In addition, the system of identity numbers was well established and easy to use from the beginning. For the Nordic countries, it was thus possible to use simple methods to produce register-based statistics on persons.

The situation in many countries is quite different from that of the Nordic countries. If the quality of the civil register is not sufficient for statistical purposes, many registers must be combined to obtain acceptable coverage. Residential addresses in the civil register may be of low quality, which means that address information from other sources must be used for the statistical population register. Different parallel systems for identities may be in use, and duplicates can be created when a person registers a second time and receives a new identity number.

Births, deaths, immigration, marriages, and divorces are registered in the civil register. In some countries, emigration, immigration, and migration within the country are not registered, which gives rise to serious statistical problems. Regional statistics will be difficult if migration within the country is not included in the national registration system.

## Emigration and Immigration

The national statistical office of Colombia (DANE<sup>5</sup>) has developed a method to classify persons as emigrants or immigrants. Persons that leave or enter Colombia have their passports scanned. They are classified as persons that have permanently left or permanently entered the country based on this administrative data for a period of at least 12 months. If these data are combined with identity numbers, they can be used to update the statistical population register. Our recommendation is that this method should also be used in other countries.<sup>6</sup> In the end, the best solution would be for each country to update its civil register with information on emigration and immigration.

## Changing Registration Habits in the Country

There have been discussions in Ecuador about how the national registration can be improved. Two alternatives are mentioned:

- The registrar's office becomes responsible for updating residential addresses and information regarding household composition.
- The municipalities will be responsible for creating and maintaining municipal registers with updated residential addresses and household composition.

Nevertheless, it will take time to implement registers handled by municipalities. Until then, the registrar's office will be responsible for updating addresses in the civil register and the Social Security Institute will be responsible for updating household composition.

A third alternative was discussed during our visit in 2015. INEC, Ecuador's national statistical office, could be responsible for an internet platform where institutions such as the Ministry of Education, universities, the Social Security Institute, the Tax Institute, and others would be responsible for updating information about the population. When citizens need service from these institutions, they must provide updated information through the internet platform, such as where they live and with whom.

Our recommendations are based on our experiences from Sweden. Instead of creating parallel structures, each working with data that are relevant for the national registration system, the registrar's office should have full responsi-

---

5 Departamento Administrativo Nacional de Estadística.

6 In addition, Sweden could use the method; there is overcoverage in the Swedish Population Register that can be reduced with this method.

bility. Municipalities and other authorities should cooperate and support the registrar's office in this work. It should be mandatory for these authorities to submit all information and corrections required by the registrar's office in their work of maintaining and updating the civil register. In Sweden, about 1.2 million corrections are submitted to the registrar's office in this way each year.

The national statistical office must participate in this work of improving the registration system, and it can contribute by measuring quality and quality improvements. However, the principle of one-way traffic must always be followed: microdata can go from administrative authorities to the statistical office but never in the reverse direction.

### A Register of Electricity Meters Could Be Used as a Frame

In Colombia and Ecuador, there are discussions about whether administrative data from electricity companies regarding users' billing addresses could be used for statistical purposes. The electricity meters are georeferenced, and this information could be of great help in linking persons to buildings.

If almost all households are electricity users, then the electricity companies' registers of users could be used as a household frame. This frame could be used to generate register-based sample surveys, which could be a cost-efficient alternative to area frame-based sampling. The electricity meter register could also be used to replace the housing part of the population and housing census with a mail questionnaire sent to all households or a sample of households.

### Data Delivery

INEC in Ecuador received the first delivery of data from the civil register with the current stock of the population as of December 31, 2014, with quarterly updates planned. The delivery consisted of two registers, one with persons that had registered and received a *cédula* number and another consisting of persons with a birth certificate who have not requested a *cédula* number.

Delivering data to INEC is a new task for the registrar's office, and the delivery methods must be improved. There was only one record for each *cédula* number, and we suspect that only the active population as of December 31 was delivered. This means that it is not possible to create retroactive population statistics.

INEC carefully edited the data from the civil register and reported the error

structure to staff from the registrar's office. This was their first experience of quality feedback from outside experts.

When INEC received the first update after the first quarter of 2015, another problem arose: the registrar's office could only send the entire population again, consisting of some 18 million records, not just the changed records. These experiences show that this part of the national statistical system must be given full attention. Delivering data to the statistical office is a new task for the administrative authorities. It is important to understand what kind of data is needed and how it should be processed and selected for delivery, and how the administrative authority should use the new quality feedback.

## 5.2 How to Find Duplicates

At INEC, the problem of duplicates in the Population Register is considered an important one. Duplicates can be found by comparing names, birth dates, and birthplaces. For example, the same person has registered twice and has two *cédula* numbers or has one registration with *cédula* number and another one without. INEC was aware that the registers it received from the registrar's office contained duplicates. How can these duplicates be found and how should they be handled? We suggest the solution described below.

Population registers can be arduous to process. INEC works with registers consisting of about 18 million records. If 18 million records are compared with all other records, this would involve about 3.24 billion comparisons to be done by computer. To make data processing easier, we decided not to use record linkage. Instead, our method is based on sorting records and comparing only consecutive records.

As shown in **Table 3**, "code Nombre" contains codes for first names; "code Paterno" and "code Materno" are codes for the first and second family names. In the table, real names have been replaced with numerical codes:<sup>7</sup> Maria could get code 1076, Gonzales could get code 1018, and so forth. There are six identifying variables with names and dates of birth. In a real application, place of birth should also be included.

---

7 As the example is based on real data, the names have been changed in this paper.

**Table 3.** Searching for Duplicates by Sorting by Identities

code Nombre	code Paterno	code Materno	YYYY	MM	DD	combined code	equal
...	...	...	...	...	...	...	...
1076	1018	1751	1986	6	10	1076.1018175119860610	0
1076	1451	396	1994	3	9	1076.1451039619940309	0
1077	426	880	1982	4	1	1077.0426088019820401	0
1078	274	1001	1978	3	25	1078.0274100119780325	0
1079	1760	1779	1991	12	3	1079.1760177919911203	0
1079	1854	23	1991	11	15	1079.1854002319911115	0
1079	1854	23	1991	11	15	1079.1854002319911115	1
1080	2198	500	1986	12	5	1080.2198050019861205	0
1081	1631	2337	1988	1	7	1081.1631233719880107	0
...	...	...	...	...	...	...	...

The strictest rule for duplicates is that if all six variables are equal, then two records are considered to be duplicates. We start the search for duplicates with the strictest rule:

- With the six variables, a combined code is created by concatenating identity variables.
- All records are sorted by the combined code.
- Subsequent records are compared: if two subsequent records have equal combined codes, we have found one duplicate.
- The two duplicate records are removed for special treatment. In INEC's case, these two records have different *cédula* numbers; other variables can also differ, such as marital status. Then the duplicate may represent information from two different points in time for the same person. These two records with different *cédula* numbers will receive the same anonymous record number in the identity database.
- If we accept one error or difference among Paterno, Materno, year, month, and day of birth, we can search for more duplicates in a similar manner.
- Small differences in Nombre can be handled by computing a similarity index. As an example, the first names Ester and Esther could have an index close to 100 percent. If the similarity is above a certain limit, such as 80 percent, and the other identifying variables are equal, then the records are treated as duplicates.
- Duplicates defined according to a specified similarity rule will be found in this way. The information from the two records belonging to the same duplicate can be combined through the same anonymous record number. The two records can represent information for the same person from different sources or points in time.

### 5.3 Undercoverage and Incorrect Residential Addresses<sup>8</sup>

Coverage errors are a problem that should receive more attention. There is a long tradition of working with sampling and nonresponse errors. The rising nonresponse rates in Statistics Sweden's surveys have recently been given a lot of attention in the Swedish media. However, coverage problems have never been discussed in the media. The Population Register used by Statistics Sweden has an overcoverage and undercoverage of between 0.5 and 1 percent for both.

The Business Register used for yearly economic statistics had, for the year 2004, overcoverage of about 8 percent and undercoverage of about 28 percent regarding the number of legal units and about 2 percent regarding total turnover (Wallgren and Wallgren, 2014: 139-140, 222). These coverage errors are also present in all sample surveys that use these registers as sampling frames.

Even if the coverage problems are considered small at the total level, the errors can be highly selective. An example is the overcoverage in the Swedish population register, which generates serious problems for statistics by country of birth. The problem has not yet been tackled.

In a country such as Sweden, with many good administrative registers available, these coverage errors can be reduced by using more sources and by improving statistical methods. In developing countries that have recently started to use administrative registers, however, we suggest the use of dual frames to estimate coverage errors—that is, combining registers with sample surveys based on area frames. The dual frame approach for sample surveys is discussed by Ferraz (2015) and Carfagna and Carfagna (2010). We will also use the calibration technique developed by Deville and Särndal (1992) to correct estimates for coverage errors.

Coverage errors in censuses are often estimated by post-enumeration surveys. The scope and methods used for these surveys differ from the methods discussed here. In a post-enumeration survey, repeated measurements are made using the same frame as in the census. The measurement errors found are used to estimate undercoverage errors in the census. A recent paper by da Silva, de Freitas, and Pessoa (2015) describes how this method was used in the Brazilian Census 2010.

---

8 This section is based on Wallgren and Wallgren (2015b).

**Figure 6** in **Section 2.4** describes the different parts of the population that are included in the register or excluded. **Figure 7** illustrates the three different kinds of surveys that are discussed here: the register survey, the register-based sample survey, and the area frame-based sample survey. Ferraz (2015) considers the case when a list-based or register-based<sup>9</sup> sample and an area frame-based sample containing the same variables are combined during the estimation stage. All variables are collected for both samples.

Here, we consider a different case that is important for countries that have started using administrative registers to create statistical registers. Coverage errors and problems with outdated residential addresses may make it necessary to combine a statistical register with an area-based sample. The statistics produced are mainly based on the register or on sample surveys based on frames created with the register. With the register, it is desirable to produce estimates by municipality. The area-based sample is used to estimate coverage errors at national and regional levels and for different categories of municipalities and persons or households. It can also be possible to adjust or correct for coverage errors with calibration conditions based on some variables in the register.

Assume that we want to produce regional statistics with the Population Register created by a national statistical office. **Table 4** lists some of the variables in the register. The municipalities have been classified in different categories based on degree of urbanization and perhaps the relative size of an indigenous population. We assume that these variables can influence the size of the coverage errors for a municipality. It may be possible to classify households or persons in different categories with different propensities of moving to another residential address. Age, level of education, and economic activity are variables that can be used here.

---

9 In this paper, we prefer to use the term “register” instead of “list.”

**Table 4.** Variables in the Three Different Surveys

Variables in the register	Variables in the area frame-based sample	Variables in the register-based sample
Identity number, Idnr	If the unit is not in the register: Idnr = missing	Identity number, Idnr
Residential address	Other variables: same as in the register	Other variables: same as in the register and statistical variables for the survey
From date of residence	Possible to add statistical variables	
District		
Municipality		
Category of municipality		
Sex		
Age		
Level of education		
Economic activity		
Category of person or household		

For an area sample, interviewers collect the same variables as in the register. We will find persons in the sample that are not in the register or have different addresses in the register. We will find undercoverage and overcoverage at different regional levels: incorrect municipality, incorrect district, and persons excluded from the register.

### Estimation of Coverage Errors

A number of important quality measures can be calculated based on comparisons between the area sample and the register: percentages of persons with incorrect residential address, delay in reporting a new address, and undercoverage rates at regional and national levels. The long-term work with improvements of the national registration system should be monitored with these measures.

**Table 5** illustrates different kinds of coverage errors. At the national level, the area sample gives an estimated undercoverage of 72,000 persons in the register. The area sample estimate is that 80,000 persons are actually living in district A, but they are registered as living in district B. These 80,000 persons are an estimated undercoverage of the subpopulation living in district A and at the same time an estimated overcoverage of the subpopulation living in district B. Persons who have left the country are another category of overcoverage that cannot be detected with the methods discussed here.

**Table 5.** Examples of Coverage Errors, Area Sample Estimates of the Population by District

		District According to Register Data					
	District	A	B	C	Not in register	All	Undercoverage
District According to Area Sample	A	2,870,000	80,000	110,000	19,000	<b>3,079,000</b>	209,000
	B	16,000	3,110,000	40,000	27,000	<b>3,193,000</b>	83,000
	C	2,000	10,000	1,530,000	26,000	<b>1,568,000</b>	38,000
	All	<b>2,888,000</b>	<b>3,200,000</b>	<b>1,680,000</b>	72,000	<b>7,840,000</b>	330,000
	Overcoverage	18,000	90,000	150,000			

**Table 6.** Uncorrected and Corrected Estimates

District	Original estimates	Corrected estimates
A	2,888,000	3,079,000
B	3,200,000	3,193,000
C	1,680,000	1,568,000
<b>All</b>	<b>7,768,000</b>	<b>7,840,000</b>

In **Tables 5 and 6**, the original register estimates of the population in the districts can be compared with the area sample estimates. In the register, *district A* is underestimated, and *districts B and C* are overestimated. We can compare register estimates and area sample estimates for other categories in a similar way.

## Dual Frame Estimation

We assume that the area sample can give us reliable estimates at the regional level and that municipalities define the desired regional level for the estimates based on the register. Estimation with the dual frame approach has some shortcomings, which will make the dual frame estimators discussed in Ferraz (2015) difficult to use for the case we consider here:

- We do not think that the area sample can be designed so that dual frame estimates for municipalities can be produced. Cost limits will make this impossible.
- There will be inconsistencies between register data and measurements made by the interviewers. For example, some persons will say that they have completed secondary school but, according to school register data, they have not.

- Regional register data and register-based sample surveys will be influenced by overcoverage errors. The dual frame estimators will only correct for undercoverage. Overcoverage in the register will be detected in the area sample when the interviewers meet persons that have moved to addresses in the sample. Regarding persons that have moved from addresses in the area sample, the interviewers will understand that these addresses are incorrect as the residential address of these persons, but it will be difficult to determine if the district and/or municipality is correct or not.

Our conclusion is that the dual frame methods discussed by Ferraz (2015) are not suitable when there is both overcoverage and undercoverage for important domains of study. Errors in the register (list frame) regarding the classification variables used to define the domains of study will generate these kinds of coverage errors. Location of residence of persons and establishments are examples of such classification variables. We need a different kind of dual frame approach if we want to produce statistics based on the Population Register and correct for coverage errors.

### Adjusting for Coverage Errors through Calibration of Weights

The alternative dual frame approach discussed in this section uses methods that are similar to the calibration method we use to correct for nonresponse in sample surveys. The area sample discussed in the previous sections can be used in two ways:

1. To give estimates of errors regarding coverage and errors in basic classification variables in the register as described in the previous section on dual frame estimation.
2. To give estimates that can be used as calibration conditions for adjusting or correcting for coverage errors.

**Table 7** shows the structure of the Population Register. The variables in *columns 2-6* can be used to define calibration conditions if we want to adjust for coverage errors. If we choose to use only district and information on the number of persons from each district as calibration conditions, the original weights  $d_i$  in *column 7* will be transformed into the calibrated weights  $w_i$  in *column 11*.

**Table 7.** Seven Records from the Population Register

Identity number	Sex	Age	Municipality	Category	District	$d_i$	$x_1$	$x_2$	$x_3$	$w_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
100002	2	99	1130	2	B	1	0	1	0	0.998
100003	2	33	1224	4	C	1	0	0	1	0.933
100012	1	73	1240	2	C	1	0	0	1	0.933
100013	1	19	1134	3	B	1	0	1	0	0.998
100019	1	64	1243	4	C	1	0	0	1	0.933
100023	1	24	1040	2	A	1	1	0	0	1.066
100026	1	72	1025	1	A	1	1	0	0	1.066
...	...	...	...	...	...	...	...	...	...	...

If we sum *column 7* by *district*, we will get the original uncorrected estimates in **Table 6**. If we sum *column 11* by *district*, we will get the corrected estimates in **Table 6**.

In this case, it is easy to calculate the calibrated weights. Weights  $w_i$  for the records with *District A* in the register will be  $3079/2888 = 1.066$  and the weights for the other districts are calculated in a similar way. With columns 8-10, the same weights can be derived with matrix algebra. The data from the area sample should be analyzed, and differences regarding coverage errors for different categories of sex, age, category of municipality, and district can be used to build a better model that can be used to adjust or correct the estimates for coverage errors. This methodology will be very similar to the work that is done to adjust for nonresponse in sample surveys.

If there are more statistical registers with data on, for example, education, employment, or income, these registers should use the population defined by the population register as their register populations. The calibrated weights  $w_i$  should be used for estimation in these registers as well. These registers will also be corrected for coverage errors in this way. These corrections are consistent, and the statistics produced with all these registers will be coherent.

### Register-Based Sample Surveys

Sample surveys that use the Population Register as a sampling frame should also be corrected for the coverage errors in the population register. When adjusting the weights used for producing estimates for the sample survey, the estimates are corrected for both nonresponse and coverage errors if register

totals that have been corrected for coverage errors are used. In this way, we combine the information in the area sample with the information in the register-based sample to obtain estimates that have been corrected or adjusted for coverage errors. We have linked an area frame with a list frame with the terms used in Ferraz (2015) to produce the best possible estimates. However, the statistical problem differs from the problem studied by Ferraz, as we consider register-based statistics and use the area sample to correct the register-based estimates.

## Discussion

Many countries have population registers, business registers, and farm registers that have both overcoverage and undercoverage problems. Basic register information such as residential address or economic activity can be old and incorrect. These quality factors should be monitored by area sampling. The quality of the register-based estimates can also be improved by calibrating weights. Variables in the registers should be used to define calibration conditions that will reduce the coverage errors. In the end, the national registration systems should also be improved so that the coverage problems are gradually reduced.

Presently, we publish estimates from sample surveys with nonresponse rates that are 40 percent or sometimes more. We can do this because we trust our methods for adjusting estimates for nonresponse with the information in the registers we use as frames. Could we also adjust register-based estimates when we have 40 percent undercoverage? For us, this idea is new and surprising. As statisticians from Sweden, we are used to a situation where undercoverage is at most 0.5–2 percent. This undercoverage is considered so small that it can be ignored. First, we want to make clear that we now understand that undercoverage errors can be very selective and that overcoverage or undercoverage between 0.5 and 2 percent should not be ignored, as serious errors can be generated. We also remember that when we studied statistics as young students, we were taught that if the nonresponse rate was more than 5 percent, the sample survey was worthless!

Our conclusion is that attitudes toward errors in surveys change over time. If we trust our methods for adjusting estimates for nonresponse and/or coverage errors, we can make corrections and publish. Correcting estimates for nonresponse and correcting for coverage errors with the method we propose here requires very similar methods. Therefore, our recommendations are as follows:

- To combine register data with data from an area frame-based sample,
- To analyze the coverage problems and define suitable calibration conditions, and
- To correct estimates and publish if you have confidence in your correction method.

## 6 Conclusions

We return to the conclusions in **Section 2.5** and add some comments: If administrative registers are used for statistics production, costs can be saved and quality can be improved. Censuses can be replaced by register surveys, and new kinds of regional and longitudinal statistics can be developed. The national system can be changed into an integrated statistical system with consistent and coherent statistical registers and sample surveys. During the transition period, it may be necessary to combine register-based statistics with estimates from area samples.

If this vision is to become reality, permanent cooperation will be required between ministries, national statistical offices, and the authorities. Both senior management and staff working with administrative registers must be involved in this cooperation. In addition, the work within the national statistical office must be organized in a new way. The production process model in **Section 4.1** shows how teamwork with different register units can be organized. Clear responsibilities must be defined to reduce duplication of effort and improve consistency. Managers at the statistical office should be involved in this coordination.

Data can be used efficiently in an integrated register system, which saves costs, and populations and variables can be compared, which provides opportunities to find and reduce errors. We have found that this seemingly simple principle is very difficult to implement. People working with different surveys have different interest areas and different competencies. In addition, more than one manager will be involved, each with his/her own budget. The monthly enterprise sample survey and the yearly register with enterprise data at Statistics Sweden can illustrate this. In November of year  $t$ , the frame population for the monthly enterprise sample survey regarding January of year  $t+1$  is defined. Small enterprises are not included in this frame. For February–December of year  $t+1$ , similar monthly enterprise sample surveys are carried out.

During September of year  $t+2$ , a register unit at Statistics Sweden receives income declarations for the year  $t+1$  from all enterprises. Then the populations and variables in the 12 monthly enterprise sample surveys can be compared with the yearly enterprise register. The frame errors in the sample surveys may be serious, and it would be important to monitor these errors and try to find methods to reduce them. However, staff members at the register unit do not think they can use the monthly enterprise sample survey data, and staff members at the sample survey unit do not think they can use the enterprise

register data, so no one analyzes the quality by comparing these two sets of data. As a result, the errors go unnoticed and remain year after year.<sup>10</sup> The one-survey-at-a-time paradigm prevails, and the register system is not used to its full potential.

These experiences explain why we recommend that a central register unit should be organized. This unit should oversee the production system and measure and reduce errors and inconsistencies. Leading statistical offices in such countries as the United States and Canada still have mainly traditional statistical systems, where the one-survey-at-a-time paradigm is relevant. These institutes influence statistical thinking in other countries, which can explain why methodologists and managers in many countries have not yet understood the systems approach.

---

<sup>10</sup> Wallgren and Wallgren (2014: 130–140) discusses the coverage errors that can be detected and corrected.

## References

- Carfagna, E. and A. Carfagna. 2010. Alternative Sampling Frames and Administrative Data: Which Is the Best Data Source for Agricultural Statistics? In R. Benedetti, M. Bee, G. Espa, and F. Piersimoni (eds), *Agricultural Survey Methods*. Chichester, United Kingdom: John Wiley & Sons, Ltd.
- da Silva, A. D., M. de Freitas, and D. Pessoa. 2015. "Assessing Coverage of the 2010 Brazilian Census." *Statistical Journal of the IAOS* 31: 215–25.
- Deville, J. and C.-E. Särndal. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association* 87: 376–82.
- ECLAC (Economic Commission for Latin America and the Caribbean). 2020. Generic Law on Official Statistics for Latin America. Santiago, Chile: ECLAC. Available at: [https://repositorio.cepal.org/bitstream/handle/11362/45253/1/S2000046\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/45253/1/S2000046_es.pdf).
- FAO (Food and Agricultural Organization). 2010. Global Strategy to Improve Agricultural and Rural Statistics. Report Number 56719-GLB. Washington, DC: World Bank.
- Ferraz, C. 2015. Linking Area and List Frames in Agricultural Surveys. Paper written for FAO. Available at <http://www.fao.org/3/ca6438en/ca6438en.pdf>
- Laitila, T., A. Wallgren, and B. Wallgren. 2012. Quality Assessment of Administrative Data: Data Source Quality. Part 2 of third deliverable of work package 4 of the BLUE Enterprise and Trade Statistics Project. Available from the authors at [ba.statistik@telia.com](mailto:ba.statistik@telia.com).
- Statistics Canada. 2009. *Statistics Canada Quality Guidelines*, 5<sup>th</sup> edition.
- Stender P., T. Thorsen, and H. Andersen. 2015. Microdata Integration for Labour Market Account. *Statistical Journal of the IAOS* 31: 249–256.
- Wallgren A. and B. Wallgren. 2012. Estadísticas basadas en registros. Aprovechamiento estadístico de datos administrativos. Distrito Federal, Mexico: INEGI.

- . 2014. *Register-Based Statistics: Statistical Methods for Administrative Data*, 2<sup>nd</sup> edition. Chichester, United Kingdom: John Wiley & Sons, Ltd.
- . 2015a. "Quality Assessment in Production Systems with Registers and Sample Surveys." *Statistical Journal of the IAOS* 31: 241-47.
- . 2015b. Frames and Populations in a Register-Based Statistical System. Paper presented at ISI 2015, Rio de Janeiro, July 26-31.
- . 2016. Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos. Aguascalientes, Mexico: INEGI. Available at: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825086206>.
- . 2021 (forthcoming). *Register-Based Statistics: Registers and the National Statistical System*, 3<sup>rd</sup> edition. Chichester, United Kingdom: John Wiley & Sons, Ltd.



