



## Estimating Frame Coverage Issues in Agricultural Surveys

Anders Wallgren and Britt Wallgren  
BA Statistiksystem AB, formerly of Statistics Sweden  
Norrhagevägen 30  
SE-719 93 Vintrosa, Sweden  
ba.statistik@telia.com

### ABSTRACT

The Global Strategy by the World Bank (2011) is a strategy for improving agricultural and rural statistics in developing countries. Increased use of administrative registers for statistical purposes is one of the strategies that are recommended in the Global Strategy. When developing countries start to use administrative registers for statistics production, coverage problems will be the main quality issue due to the fact that many administrative systems in developing countries do not function properly today.

The transition period from a statistical system without registers to a system with registers will take many years, perhaps 10 years at least. During this period administrative systems will be improved and more administrative registers will be available and also the quality of administrative data will gradually become better. Many countries want to replace the Population and Household Census with statistical registers. The preconditions for agricultural surveys will then be completely different as the Population Census can no longer be used as a frame for agricultural surveys. An overview of how this transition should be handled to develop agricultural and rural statistics is given in Wallgren and Wallgren (2016b). It should be noted that *agricultural statistics* and *rural statistics* require different kinds of frames and register populations with different coverage problems.

According to the Global Strategy, all agricultural sample surveys should use an area-based master sampling frame. When agricultural registers are being created in the future, this sampling frame will get an alternative in the statistical Farm Register. This Farm Register can be used both as sampling frame with many potential auxiliary variables and also as register population for the statistical registers with data on agriculture. When the Farm Register is used in this way, sampling-based and register-based agricultural statistics can be consistent. The area-based master sampling frame has theoretically no coverage errors, but the Farm Register and all surveys based on the Farm Register will have problems both due to undercoverage and overcoverage. Until the administrative systems have become efficient, these coverage problems can be substantial.

The Global Strategy also states that agricultural statistics should be based on an *integrated system* of surveys. This means that frame populations and register populations should be consistent

so that estimates from different surveys will be coherent. This raises the question on how we should combine area based sample surveys with register surveys – a *dual or multiple frame approach* can be helpful to reduce coverage errors and improve consistence. Finally, the Global Strategy says that agricultural statistics should be an *integrated part* of the national statistical system. This means that frames and register populations regarding nonagricultural statistics must fit with the agricultural frames and register populations so that all statistics produced will be consistent and coherent.

The conclusion of this is that the Global Strategy raises many important and difficult issues regarding coverage and coverage errors of frames and register populations in the national statistical system. In this paper we explore these issues. Our field of concern consists of coverage issues in register surveys and register-based sample surveys. Multiple frame methods should always be used when statistical registers are created – you should always use *all* relevant sources. We present also estimation methods based on calibration of weights, but the most important methods for reducing coverage errors consist of hard and long-term work with improving administrative systems. The methods we present are adopted to countries in Latin America and the Caribbean, as we presently are working together with a number of countries in that region.

**Keywords:** Frame, register population, coverage error, dual frames, multiple frames, calibrated weights, integrated statistical system.

## 1. Register surveys and estimation methods

We distinguish between the following kinds of statistical surveys (we omit censuses here). Frame errors or coverage errors are generated in different ways for these kinds of survey:

1. Area frame based sample survey. Theoretically no frame errors.
2. List-based sample survey. The sampling frame is based on a list that is not updated, for example an old census. Frame errors increase when the list becomes older and older.
3. Register-based sample survey. The sampling frame is based on one or more administrative registers that are continuously being updated (every day, month, quarter or year). Frame errors depend on the administrative systems that are used to update the registers and the register-statistical methods that are used when the frame is created.
4. Register survey. A statistical register is created with available statistical and administrative registers for a specific statistical purpose. All relevant registers should be used. Here we should not use terms as *frame*, *frame population* or *frame errors*, these terms are terms from the sampling theory and they are not adequate here. Instead we should use the terms *register population* and *coverage errors*. Coverage errors in register surveys depend on the registers that are used to create the new statistical register and the register-statistical methods that are used when the new register is created.

With a dual frame approach or a multiple frame approach more than one frame is used with the intentions to reduce frame errors (coverage errors). The traditional approach is to combine an area frame with a list-based frame (1 and 2 above), but we will also consider combinations of an area frame and a register-based frame (1 and 3 above) and also combinations of an area sample and a register survey (1 and 4 above). It should be noted, that even if the register-based sample survey (3 above) or the register survey (4 above) are not combined with area samples, these methods are actually based on a multiple frame approach as many registers are combined. This is a standard rule in register-based statistics: to improve quality regarding content, coverage and variable values, *all relevant sources* should be combined when a new register is created.

The estimation methods used for the sample surveys 1-3 above are the usual sampling theory estimates, but the estimation methods used for register surveys require some explanation. For a sample survey we decide on the sampling design, how auxiliary variables should be used and how we will adjust for nonresponse. We derive sampling weights and calibrated weights and express the estimators with formulas.

Also for register surveys we should use the term *estimator*. It is equally important when we work with register surveys to distinguish between the actual values in the target population and the estimates produced with the statistical register. But where are the estimators? As a rule, there are no formulas that can express the estimators in a register survey, instead it is *the way that the statistical register has been created* that determines the values of the estimates in a register survey.

In Chart 1 the estimates are based on four registers, which in turn are based on other registers. Columns 1 and 2 are based on the Business Register, columns 3 and 4 on the Population Register (the LFS uses a frame that has been created with the Population Register). Also, all four surveys use the variable *economic activity* from the Business Register. The four surveys give estimates describing the population of employees during November 2004 in Sweden<sup>1</sup>. The estimates differ considerably<sup>2</sup>; different sources have been used and the four data sets have been created in different ways. We consider these *choices of sources and methods* for creating registers as *estimation methods* as these choices determine the estimated values and the quality of the estimates.

**Chart 1. Employees by economic activity in Sweden, November 2004, thousands**

Economic activity	Business Register		Employment Register	Labor Force Survey	Error margin
	Enterprises (1)	Local units (2)			
Mining, quarrying, manufacturing	688	636	717	640	23
Wholesale and retail trade	456	453	484	456	20
Government	139	215	239	230	15
Education	382	408	431	462	20
Health and social work	836	684	675	675	24
Other service activities	142	163	175	168	13
...	...	...	...	...	...
<b>Total</b>	<b>3 763</b>	<b>3 763</b>	<b>3 924</b>	<b>3 778</b>	<b>43</b>

In Chart 1 all surveys have coverage problems. We have analyzed coverage and found undercoverage and overcoverage errors in all of them. The variable *economic activity* is also treated differently, so that more inconsistencies are created. The conclusion is that these four surveys are not an *integrated system* of surveys. Other important parts of Statistics Sweden's production system are perfectly integrated with consistent populations and variables.

## 2. Improving the national statistical system

Since 2001 we have visited a number of countries in Latin America and the Caribbean. The countries in the region want to start to use administrative registers for statistics production. In Wallgren and Wallgren (2016b) we discuss what strategy and methods that should be adopted by

<sup>1</sup> This is a real example with Swedish labor market statistics. An agricultural example from a developing country could show similar problems, e.g. employed in agriculture by type of farm/holding (rice producer, coffee producer etc.) from three registers and one agricultural sample survey.

<sup>2</sup> Columns 1 and 2 have low quality, column 3 contains estimates from the best of these surveys. Coding of economic activity does not function well in the Labor Force Survey, that is why we prefer the Employment Register in column 3.

developing countries that want to modernize their statistical systems for agricultural statistics and start using administrative registers. Here we want to highlight the coverage issues that are important in this context.

To analyze coverage of frames and registers it is necessary to have microdata with identity numbers. In Ecuador a test with a traditional census including identity numbers was done on the Galapagos Islands during 2015. This census was compared with available administrative registers for Galapagos. To ask for the identity numbers of the persons in the census gave no problems and the identity numbers had good quality so that deterministic record linkage could be done with the population register for Galapagos<sup>3</sup>.

**Chart 2.** Coverage errors in the population register for Galapagos 2015 (cannot yet be published)

Test census 2015 for Galapagos	Population Register for Galapagos 2015
Population size, N(census)	Population size, N(register)
Deterministic matching with register Matching key: Identity number	Deterministic matching with census Matching key: Identity number
Probabilistic matching with register Matching key: Names, 85 % similarity	Probabilistic matching with census Matching key: Names, 85 % similarity
Undercoverage in the register: persons are in the census, but not in the register	Overcoverage in the register: persons are in the register, but not in the census

Some important conclusions can be drawn from the results in Chart 2. Some conclusions are general and some are specific for this example.

1. The quality of administrative registers must always be analyzed. In this case, administrative registers for a small region are compared with an area frame based *census* for the same region. Another method is to compare a national register with an area frame based *sample survey*.
2. When a system with national identity numbers has been established, *probabilistic record linkage* becomes a redundant method. *Deterministic record linkage* will become the method that will be used, and this will open new possibilities for the national system. In the Nordic countries many record linkage operations are required to create many of the statistical registers, e.g. longitudinal registers may require more than 100 record linkage operations, and this requires that systems with identity numbers have been established. The example in Chart 2 shows that Ecuador’s identity numbers can be used for deterministic record linkage.
3. The population register for Galapagos cannot be used for statistical purposes – the coverage errors are very serious. It is clear that people leave or go to Galapagos without reporting to the office that is responsible for the register. With the present system for national registration in Ecuador, migration is not recorded with sufficient quality. Regional statistics thus becomes impossible.

The main conclusion of the results in Chart 2 is that the national registration system in Ecuador must be improved, otherwise a modernization of the statistical system will be impossible. Many countries in Latin America face the same problem.

<sup>3</sup> Based on information from our colleagues at Unidad de Metodología de Registros Administrativos, Dirección de Registros Administrativos, Instituto Nacional de Estadística y Censos (INEC), Ecuador.

## 2.1 Improving coverage – improving the national administrative systems

The most important method to reduce coverage errors in registers is to improve the administrative system that generates the data. Special estimation methods as multiple frame estimation can relieve the symptoms but not cure the disease. The Population Registers in many Latin American countries have coverage errors of the kind that is described in Chart 2. The Cadasters or Real Estate Registers are handled by the municipalities and in poor and rural parts of the country the Cadaster can be missing or in a very bad condition. The Business Registers, Farm Registers and Social Insurance Registers only include the formal part of the economy. Persons and households are not georeferenced today, but must be georeferenced so that it will be possible to produce regional and rural population statistics. Jobs and establishments must also be georeferenced so that regional labor market statistics can be produced.

If it is decided that a country will try to modernize the statistical system, then these problems must be tackled. This is a long term project and requires cooperation between politicians, the national statistical institute and ministries and other administrative authorities. The Nordic countries have all gone through this kind of process – legislation and the national registration system were changed and hard work was required. Today we find that some countries in Latin America has started this process. Ecuador is a leading country that has started to improve the national registration system.

## 2.2 Frames and the increasing use of administrative registers

When a Population Register, a Business Register and a Farm Register have been established and are used for statistics production, area frame based sampling will gradually be replaced by register-based sampling. In the Nordic countries, area sampling was abandoned in the 70s. Until the registers have sufficient coverage, it will be necessary to continue with area based sampling.

The transition from area sampling into register-based sampling is connected with the following important changes:

1. National identity numbers are used in all administrative systems, all public systems and also some systems in the private sector. There should be only *one* system of identity numbers for persons, one system for enterprises, etc.
2. Identity numbers are used in all surveys and registers. This makes integration of microdata from different sources possible.
3. There will be a shift from household surveys into surveys of persons; from establishment surveys into enterprise surveys. The reason behind this is that persons and enterprises/legal units are the units that are used in administrative systems.
4. The NSIs must create and maintain statistical registers with establishments and holdings as there will be no administrative register with these statistical units.

When the Population Register, Business Register or Farm Register are used for creating frames, these registers should contain important variables and classifications that can be used in the sample surveys. Age, sex, residence, economic activity etc. are examples of register variables that can be included in the frames. In a statistical system that is based on registers, rural statistics requires access to a Population Register with a classification of persons' residence as rural or non-rural.

When a Farm Register has been established, the holdings belonging to family farms will be identified with the holders' personal identity numbers. Commercial farms will be identified with business identity numbers. Administrative sources with agricultural data can thus be linked with the

Farm Register. As the Farm Register will be used to create frames, also all sample surveys will be linked with the Farm Register. By using microdata with identity numbers, the coverage of frames and registers can be checked and consistence regarding populations can be improved. In this way, the agricultural surveys will be an *integrated system* according to the ideas in the Global Strategy.

If the national statistical system includes a Population Register and a Business Register, then the units in the agricultural statistical system can be linked to the national system. If inconsistencies are checked and reduced, the agricultural system will be an integrated part of the national system according to the ideas in the Global Strategy.

### 3. Coverage errors and dual frame estimation

In a traditional dual frame approach, one list-based sample survey is combined with one area sample and the surveys are done simultaneously and the same variables are collected in both sample surveys. The estimates are based on data from a combination of the two sample surveys.

In the dual or multiple frame applications we discuss here, one register with coverage problems is combined with one area sample. The register is created first and thereafter data is collected for the area sample. The main basis for the estimates will be the variables in the register.

We will also consider the case when one sample survey that uses a register with coverage problems as sampling frame is combined with an area sample. In the register-based sample, both register variables and variable values that have been collected in the sample survey are used.

When developing countries start to use administrative registers for statistics production they will face serious coverage problems related to the new administrative sources. There are different methods or strategies that should be used to reduce these coverage errors:

1. *Improve the administrative systems that generate the administrative registers.*

This is the most important method and efforts should always be taken to work together with politicians, ministries and other administrative authorities so that the administrative systems in the country are modernized and improved. These improvements are necessary even if the improved statistical usability is disregarded – developing countries cannot continue to have bad and inefficient administrative systems! However, this is a long term strategy and must be combined with some or all of the following methods during the transition phase from a traditional statistical system into a register-based system:

2. *Combine many registers ('frames')* with coverage problems to create one statistical register with better coverage. This is an important method that can be used at low cost. The search for sources that can be used to improve the quality of statistical registers should always go on. Even if only a small part of the register is improved it is worthwhile to use additional sources. We give here two examples of sources that can improve the quality of the statistical Population Register:

– *In the system for registration of births that is used in Latin America, the residential address of the mother is recorded. The quality is much better than the address in the civil register.*

– *In the welfare systems for support of poor households in some countries in Latin America, the residential addresses of the households that get support are updated every year.*

In the final statistical Population Register there should be technical variables describing the source used for the address in each record and information on when the address was updated. This information is necessary for quality assessment and will be used by the estimation methods described below.

3. *Evaluate the coverage of the new statistical registers that are being created.*

If identity numbers are included in all censuses, sample surveys and registers, quality and coverage can be evaluated through comparisons of sources with microdata:

- *Compare registers with area frame based samples. It will be possible to use already existing sample surveys and only add a few questions that can be compared with a small number of variables in the register for the selected units. During the transition period statistical registers as the Population Register, the Business Register and the Farm Register can be evaluated in this way until quality and coverage is sufficiently good.*
- *Compare census data with register data. If traditional censuses are done, the corresponding registers can be evaluated and improved. This was standard practice in the Nordic countries during the transition period when the traditional Population and Housing Census was replaced with statistical registers.*
- *Compare different registers with each other. You will learn a lot about quality and coverage.*

4. *Develop dual frame estimators for register-based statistics.*

In the sections that follow, we try to translate the methods used for dual frame estimation with one list-based sample (with undercoverage) and one area sample into methods that can be used in a register-based system. The idea is to adjust for coverage errors in the registers with information from an area sample with no coverage errors. We will introduce *calibrated weights* in the registers. The common approach is to use register data to calibrate the sampling weights in sample surveys with nonresponse errors with the aim to reduce these errors. Here we work in the opposite direction: we use sample survey data to calibrate weights in a register with coverage errors with the intention to reduce the coverage errors.

### **3.1 Combining an area sample and a register – the Population Register**

Residential addresses are often not updated in administrative population registers in many countries in Latin America. This means that it is difficult to use administrative population data for regional statistics. In countries with strong urbanization, this will lead to overcoverage of rural register populations and undercoverage of urban register populations. A statistical Population Register is important for rural statistics and also for agricultural statistics, e.g. it is important to know if poor farmers give up farming and move to a city.

During ISI 2015 we presented ideas on how estimates based on statistical registers could be adjusted for coverage errors with calibrated weights (Wallgren and Wallgren, 2016a). The adjustments are done in a similar way as when we adjust sample survey estimates for nonresponse errors, but here we adjust the register-based estimates with a sample survey instead of adjusting the sample-based estimates with a register. This method for adjusting register estimates for coverage errors can be used during the transition period when administrative systems are gradually being improved so that coverage errors are reduced. First, coverage errors in the statistical register should be reduced by using all relevant sources. Secondly, the coverage errors should be monitored with area based sample surveys. A third step could be to adjust the register-based estimates with calibrated weights that are based on a comparison between the area sample and the register.

In many countries in Latin America, regional register-based statistics is difficult, because *migration within the country* is not recorded in the national registration systems. To test our ideas regarding calibration of weights we have used linked microdata from Statistics Sweden's Population register for 2005 and 2015 regarding persons that belonged to the population both these years. This means that the only demographic events we study are migration within Sweden.

Coverage errors in the Population Register can be caused by different problems with the registration system and each problem should be tackled with a specific long-term project. Underreporting of births and deaths should be reduced by improving registration methods and

underreporting of immigration and emigration should be reduced through cooperation with migration authorities and improved border control. The most serious problem is migration within the country the final solution is to change the nation registration system so that it becomes mandatory to report to the register when a person changes residential address. But this requires new legislation and changed attitudes and will take time. In the short term, the methods we present here can be used where estimates are adjusted for net migration.

We assume that the Swedish register for 2005 represents a bad register from a country in Latin America that has not been updated regarding migration within the country. Data from 2005 we will call *Registro Civil not updated for migration* and data from 2015 we will call *the unknown truth today*. We will use data from Registro Civil combined with sample survey data from the register for 2015 to try to estimate the unknown truth today. Migration creates undercoverage and overcoverage in regional statistics. We can compare microdata from Registro Civil with the truth for the Swedish municipality *Åsele*. The population of *Åsele* is  $619 + 2\,081 = 2\,700$  persons according to Registro Civil but only  $2\,081 + 305 = 2\,386$  persons according to the truth. The gross error due to coverage errors is  $619 + 305 = 924$ , but the net error is only  $619 - 305 = 314$ . It is only net errors (net migration) that are important for the estimates.

Overcoverage, 23 %	619	}	} 2 700, Population according to Registro Civil
	2 081		
Undercoverage, 13 %	305	}	} 2 386, Population according to the truth

To be able to adjust for coverage errors due to net migration, we need auxiliary variables in the register that are correlated with net migration. By grouping the 290 municipalities in Sweden into 10 categories, from the category with strongest net migration *into* the municipality to the category with strongest net migration *from* the municipality. By analyzing the last two censuses and by using information in other registers that are available, countries in Latin America can find variables that can be used to create similar categories. If we return to the Swedish example, we also want to try *age* as auxiliary variable as young persons have a higher propensity to migrate than old persons.

Our first calibration exercise (number 3 in Chart 4) uses only one auxiliary variable, *category* of each person's municipality of residence. Our idea is that no special sample survey is necessary for calibration of weights in the statistical Population Register, already existing household surveys based on area sampling can be used. It is only necessary to include information of each person's identity number and municipality of residence in the survey. In calibration exercise 3 we need only 10 estimates from the sample survey and the corresponding estimates from the register. The ratios between these estimates are the weights in Chart 3.

**Chart 3. Population estimates, exercise 3**

Category	RC, register	Area sample	weight
1	351 107	317 396	0.90399
2	268 733	249 748	0.92935
3	340 143	320 935	0.94353
4	385 406	369 257	0.95810
5	577 247	558 744	0.96795
6	765 929	748 678	0.97748
7	936 860	929 009	0.99162
8	1 011 517	1 014 497	1.00295
9	2 037 960	2 080 777	1.02101
10	1 288 365	1 374 226	1.06664
<b>All</b>	<b>7 963 267</b>	<b>7 963 267</b>	

In Chart 4 the 290 estimates regarding the population of each municipality are compared with the true values.

In Exercise 1, the Registro Civil estimates without corrections or weights are compared with the true values. The mean absolute error of the estimates based on Registro Civil only is 4.4 %. The population of one municipality is over-counted with 16.8 % and 75 % of the municipality estimates have errors that are smaller than 6.4 %.

In Exercise 2, Registro Civil has been supplemented with microdata from another administrative source that is common in Latin America<sup>4</sup>. Poor people who want support go to a regional office and fill in a form where information on the person's identity number and present place of residence is included. If a person moves to another municipality, the person must update the information in the poverty register (called Sisben in Chart 4) to get continued support. As about half of the population can be included in this kind of poverty register in Latin American countries, the coverage errors in Exercise 2 are reduced with about 50 %.

In Exercise 3 we introduce weights in the version of Registro Civil that has not been updated. The 290 municipalities have been grouped into 10 categories, from category 1 with strongest net migration out from the municipalities up to category 10 with strongest net migration into the municipalities. With these categories as auxiliary variable the quality of the estimates is improved.

In Exercise 4 we also improve Registro Civil with the information in the Sisben Register. The quality of the estimates is now quite good – if Registro Civil that has not been updated during 10 years is supplemented with information on residence in the poverty register and if estimation uses weights derived from the simple model in Chart 3, then the municipality population estimates are reliable. In the Sisben part of the register, persons will have weights close to 1, but in the remaining part of the register the weights vary between 0.90 and 1.07.

**Chart 4. Reducing coverage errors in population estimates**

Calibration exercise:	Mean absolute error, %	3 <sup>rd</sup> quartile	Maximum absolute error, %
1. Registro Civil, not updated for migration, (RC)	4.4	6.4	16.8
2. RC supplemented with Poverty Register (Sisben)	2.2	3.2	9.0
3. RC, Calibrated weights, 10 categories	0.6	0.6	6.0
4. RC and Sisben, calibrated weights, 10 categories	0.4	0.5	3.3
5. RC and Sisben, cal. weights, 10 cat. and age groups	0.4	0.5	3.8

In Exercise 5 we also use *age group* as an auxiliary variable together with the categories that were used in Exercise 4. With seven age groups combined with ten categories, the area sample is used to estimate 70 parameters. In Chart 4 it is clear that the variable *age* does not improve the quality of the estimates of the municipality populations, but it is clear that the age distributions of the persons in the municipalities are improved. Persons 20-39 years with high propensity for net migration are increased by the weights for Stockholm municipality, but decreased for Åsele.

**Chart 5a. Estimates for Exercise 1 (RC) and Exercise 5**

Stockholm municipality (category 10)						Åsele municipality (category 2)					
Age	RC	Ex. 5	Truth	Error RC	Error ex.5	Age	RC	Ex. 5	Truth	Error RC	Error ex.5
10-19	11.6	10.6	10.3	1.4	0.3	10-19	10.7	11.2	11.1	-0.4	0.1
20-29	13.4	15.8	17.0	-3.6	-1.2	20-29	14.6	10.3	11.2	3.3	-0.9
30-39	15.5	17.6	17.9	-2.4	-0.3	30-39	10.4	7.8	8.6	1.8	-0.8
40-49	17.7	16.3	15.6	2.1	0.7	40-49	10.5	10.7	10.5	0.0	0.2
50-59	15.3	14.8	14.7	0.7	0.1	50-59	14.3	15.3	15.1	-0.8	0.2
60-69	12.8	12.3	12.1	0.8	0.2	60-69	16.1	18.2	17.8	-1.7	0.4
70+	13.7	12.8	12.5	1.1	0.2	70+	23.5	26.6	25.6	-2.1	0.9
All	100.0	100.0	100.0			All	100.0	100.0	100.0		
Population	670 474	717 621	718 008			Population	2 700	2 433	2 386		
Error %	-6.6	-0.1				Error %	13.2	2.0			

<sup>4</sup> We created an artificial register of this kind.

### 3.2 Combining an area sample and a register – the Business and Farm Register

Our conclusion of these estimation exercises is that it is possible to improve regional statistics based on the *Population Register* in countries in Latin America with the methods we recommend here. We have used real Swedish data with errors that we think are similar to what countries in Latin America experience. With a simple model with two auxiliary variables, statistics for small regions was improved.<sup>5</sup>

In the administrative *Business Registers* in Latin American countries, only formal enterprises are included. Only the formal sector will be found in the administrative sources and these registers must be combined with area samples and censuses that cover both the formal and informal sectors. The administrative registers can have an undercoverage of about 50 %. This means that the methods we present in Section 3.1 cannot be used here. If a *Farm Register* has been established in the way we discuss in Wallgren & Wallgren (2016b), the family farm part of the register is closely related to the Population Register. In this case, it may be possible to use the methods in Section 3.1 to adjust estimates based on the Farm Register.

### 3.3 Combining an area sample and a register-based sample

In a traditional dual frame approach, one list-based sample survey is combined with one area sample and the surveys are done simultaneously and the same variables are collected in both sample surveys. For register-based sample surveys we suggest a completely different method:

1. The area sample is only used to adjust the Population Register.
2. The register-based sample uses this adjusted Population Register as sampling frame, the weights in the Population Register are included in the data matrix for the sample.
3. The sample estimates are calculated with sampling weights that are the design weights adjusted for nonresponse, multiplied with the weights from the Population Register. The sample estimates are then adjusted both for nonresponse and coverage errors.

Sample surveys done in this way will be consistent with the Population Register. Also all other statistical registers with data on persons will use the adjusted Population Register as register population and as the same weights are used, these registers and sample surveys will be an integrated system with consistent statistics.

## REFERENCES

- Wallgren, A., Wallgren, B. (2016a) Frames and Populations in a Register-based National Statistical system. *Journal of Mathematics and Statistical Science*, 2016, pp. 208–216.
- Wallgren A., Wallgren B. (2016b) Administrative data and Agricultural Statistics – What Strategy and Methods should we adopt? Paper presented at ICAS VII, Rome October 2016.
- World Bank (2011) *Global Strategy to Improve Agricultural and Rural Statistics*, the World Bank, FAO and the United Nations.

---

<sup>5</sup> We have not mentioned or analyzed the sampling errors in the area based sample. Our reason for this is that the main source of errors in this case depends on the model that is used to create the weights. What we need is strong correlation with the category for each municipality and the real net migration of that municipality. The sampling errors in the area sample should also be considered, but depends on the survey design and sample size used in the specific country.