



**Forwarding Armenian Statistics Through Twinning**

**AM09/ENP-PCA/TP/04**

**STUDY VISIT REPORT**

on

**QUALITY MANAGEMENT**

ACTIVITY A5.2: STUDY VISIT:

**ORGANISATION OF SAMPLING ISSUES  
IN STATISTICS DENMARK**

Study visit participants  
Ms Lilit Petrosyan  
Mr Laert Harutyunyan

20-24 August 2012

Final version



National Statistical Service  
Republic of Armenia



Statistics Denmark

**Authors' names, addresses, e-mails**

Ms Lilit Petrosyan  
National Statistical Service of RA  
3, Government Building  
Tel. 00374-10- 52-32-04  
E-mail: [lpetrosyan@armstat.am](mailto:lpetrosyan@armstat.am)

Mr Laert Harutyunyan  
National Statistical Service of RA  
3, Government Building  
Tel. 00374-10- 52-42-45  
E-mail: [business\\_reg@armstat.am](mailto:business_reg@armstat.am)

## Table of contents

1. General comments	4
2. Assessment and results	4
3. Implementation of the study visit	4
4. Results of the study visit	5

Annex 1	Terms of Reference
Annex 2	Programme for the study visit
Annex 3	Persons met
Annex 4	Data provided before the study visit
Annex 5	Statistics Denmark's analysis of Labour Force data (sampling error)

## List of Abbreviations

EU	European Union
NSSRA	National Statistical Service, Republic of Armenia
RA	Republic of Armenia
SBR	Statistical Business Register
SBS	Structural Business Survey
SD	Statistics Denmark

## **1. General comments**

The study visit report was prepared as the result of the study visit on Quality Management to Statistics Denmark, organized 20-24 August 2012, that focused on sampling issues. The BC Experts would like to express their gratitude to the representatives of Statistics Denmark for providing valuable methodological information and for their kind support during the stay.

## **2. Purpose of the activity**

The main purpose of the study visit was to obtain an overview of the way sampling issues is organized in Statistics Denmark. It is expected that sample surveys will play an increasingly important role in the Armenian statistical system. Accordingly, more focus will have to be put on sample design and other optimization procedures that will be necessary to obtain a high quality even though data collection will be reduced. The experience from Denmark should serve as inspiration for this. Furthermore, other quality issues were discussed

## **3. Implementation of the study visit**

The following issues were discussed during the study visit:

- Study of the organisation and division of work between the subject matter divisions and the methodological unit of Statistics Denmark;
- Small area estimations, relevant for regionally distributed statistics;
- Introduction to sampling error estimation;
- Optimization of sampling as a means of reducing response burden;
- Examples of use of administrative registers for statistical purposes;
- Questionnaire design in Statistics Denmark and further input to the on-going process of redesigning the reporting forms to business statistics of NSSRA

On the first day of the study visit, Ms Charlotte Hansen of Statistics Denmark's International Consulting division welcomed the group and introduced the historical background of Statistics Denmark, its mission and vision, data collection, processing and dissemination procedures, main functions of the units as well as modern development trends.

Later, the head of Methodology Division Mr. P Linde presented the main functions of the division. It was mentioned, that the division was created in 2003 and initially there were only three employees. Before now the division was a part of Business Statistics with 9 employees, particularly 6 of them are involved in mathematicians' section dealing with sampling, imputation, standard errors, calibration, etc, while 3 of them work in questionnaire development section (including electronic questionnaires).

As a matter of fact except with the purpose of establishing sampling population for subject matter divisions, one section of the division works on the implementation of seasonal adjustments for all subject matter divisions. It was also mentioned that starting from 01/09/2012 the division will be reorganised and Mathematicians section will join with „Survey services” division. The role of the Methodology division was emphasized in the context of increasing quality of statistics. The important task for estimation of standard error was introduced. The practice of Statistics Denmark implementation of sample surveys was introduced in detail as well as the work processes of subject matter units and units responsible for the data collection. Other problematic issues such as compilation of sample population, estimation of existing

sample, calculation of statistical standard error, imputation as well as sample population made by NSS RA with existing gaps and possibilities for improvement were discussed with Danish colleagues.

The relevance of small area estimation in NSS RA was also discussed. It was mentioned, that it is not applicable to make small area estimations since it is mostly intended for model estimations. As for the quality assurance issues, it was mentioned that first it is possible to ensure with sampling without gaps and small sampling errors.

During study visit Mrs. A. Conrad introduced the work implemented in Statistics Denmark in the context of minimizing response burden and developing questionnaires. Particularly, it was mentioned that it was a strategic decision to start the collection of all questionnaires electronically from 2013, and it is crucial to have the user friendly questionnaires. Already existing electronic questionnaires were presented to the group.

Representatives from Business Register division of Statistics Denmark made a presentation on updating business register. In addition, the development strategy of NSS RA business register was also covered.

It was also foreseen to have a presentation on „Usage of administrative sources in agriculture statistics” during which the relevant clarifications were made on legal basis, data sources ( 5% of data is taken from surveys, while 95% of data comes from administrative sources) with the stress on the cooperation with administrative bodies, qualitative aspects of received information, and storage of the data. Later, the usage of the register with the purpose of compiling agriculture statistics was discussed with its quality assessment, and solution of the encountered issues. In addition, Danish and EU legislative documents on implementation of agricultural census for 2010 were introduced in details with relevant organisational activities for the census.

#### **4. Results of the study visits**

Study visit provided a chance to get introduced to Statistics Denmark professional experience in organizing methodological work and use them in NSS RA current work and challenges. In addition, the relevant agreement was made on formulating the strategy for the development of statistical business register. Particularly, it was decided document's scope of development trends was distinguished specifying the role of the involved units.

## **Annex 1 Terms of Reference**

### **Terms of Reference (A5.2); 20 – 24 August 2012**

<b>Component A</b>	<b>Quality Management</b>
Component B	Business Register, Structural Business Survey, and Respondent Burden
Component C	Improvement of the Exhaustiveness of GDP
Component D	Agricultural Census
Component E	Harmonized Consumer Price Index
Component F	ICT Society

#### ***Activity A5.2 Study visit: Organisation of sampling issues in Statistics Denmark***

##### **1. Purpose of activity**

The purpose of the study visit is to study the organisation of the work related to statistical methodology in Statistics Denmark. In addition, a number of specific methodological topics will be discussed of which some supplements the work within the project's Business Statistics component.

##### **2. Expected output of the activity**

The expected output and topics of the study visit are:

- Study of the organisation and division of work between the subject matter divisions and the methodological unit of Statistics Denmark;
- Workshop on so-called small area estimations, relevant for regionally distributed statistics;
- Input to the Twinning component on ICT statistics regarding regionally distributed statistics;
- Introduction to sampling error estimation;
- Optimization of sampling as a means of reducing response burden;
- Examples of use of administrative registers for statistical purposes;
- Questionnaire design in Statistics Denmark and further input to the on-going process of redesigning the reporting forms to business statistics of NSSRA.

The study visit will be documented in a report.

##### **3. Study visit participants**

Ms Lilit Petrosyan, Head of Methodological Division, NSSRA

Mr Layert Haratyunyan, Head of Business Statistics Division, NSSRA

Ms Diana Gasoyan, RTA Assistant and Interpreter

##### **Participants from Statistics Denmark (preliminary)**

Mr Lars Thygesen, Mr Jesper Ellemose Jensen, Mr Peter Linde, Mr Peter Stoltze, Ms Ashu Conrad, Mr Carsten Zornig.

## Annex 2. Programme for the mission

<b>Time</b>	<b>Event</b>	<b>Purpose / detail</b>
Monday, 27 August 10:00	Welcome	Introduction to Statistics Denmark and to the programme of the week
11:00	Workshop	Presentation of the current organisation of the methodological work within NSSRA ( <i>BC Experts</i> )
13:00	Workshop	The organisation of methodological work in Statistics Denmark. The co-operation and division of work between the subject matter divisions and the methodological unit
Tuesday, 28 August 9:00	Seminar	Introduction to sampling error estimation
13:00	Seminar (together with study visit on D.5, Agricultural Census)	Co-operation with the ministerial level, and use of administrative registers for agricultural statistics
Wednesday, 29 August 9:00	Workshop	Small area estimations – Case study: Presentation of the on-going project work on the ICT usage statistics ( <i>BC Experts</i> )
13:00	Seminar (together with study visit on D.5, Agricultural Census)	The co-operation of the Agricultural Statistics Division of Statistics Denmark with other divisions: - Business Register Division - Methodological Division - Dissemination Division - National Accounts Division
15:00	Meeting on Business Statistics	Response burden measurement: - Calculation of response burden - Updating techniques of the baseline measurement
Thursday, 30 August 9:00	Workshop	Questionnaire design in Statistics Denmark
13:00	Workshop	Presentation on and to the on-going project work on questionnaire design ( <i>BC Experts</i> ) Joint discussion and further input to the work
Friday, 31 August 9:30	Workshop	Overall implications for the project work within component A
13:00	Meeting on Business Statistics	Updating of road map and planning of necessary actions to be taken by NSSRA within the Business Statistics component
14:00	Debriefing	Other issues and evaluation of the study visit

Lunch time every day 12:30-13:00 (Statistics Denmark's canteen).

## Annex 3. Persons met

Jan Plovsing, Director General, Statistics Denmark  
Lars Thygesen, Director of Sales & Marketing, Statistics Denmark, MS Project Leader  
Jesper Ellemose Jensen, Chief Advisor, International Consulting Division  
Peter Linde, Head of Data Collection and Methodology Division

## Annex 4. Description of provided data sets before study visits

### A. ICT Usage in enterprises

#### 1 The status as of 01/08/2012

The pilot study has been executed

The questionnaire is designed according to EU regulation (model questionnaire)

Pilot study: 100 enterprises with 16 stratas, 4 size classes ( number of employees) x 4 industry groups (NACE)

Data was collected and processed.

First draft of pilot study report is currently being reviewed.

#### 2 Additional goals during the implementation of the project

Sample plan for full scale survey with 12 industry groups and 5 size classes s

The results of pilot study were discussed with external stakeholders.

The plans for future full scale survey were discussed and updated on the basis of discussions with external stakeholders.

Development strategy for ICT statistics was introduced.

#### 3 Sampling plan for full scale survey

Challenges:

1. ICT usage in 11 regions is rather different as it was supposed to be
2. The demography of enterprises is centralized in Yerevan city
3. Data distributed by regions are logical where as national average data may be illogical
4. Published data at least should have the allocation between capital and other regions
5. Very <<thin>> or empty cells in stratified population.

It is necessary to consider different sampling methods taking into consideration 1) below **mentioned description tables** 2) anonymous data at micro level that is to say according to enterprise type that will include number of enterprise, region, NACE industry, number of employees

Data should be compiled in Excel format before study visit and sent to Statistics Denmark to the following e-mail addresses: [pli@dst.dk](mailto:pli@dst.dk) and [mlu@dst.dk](mailto:mlu@dst.dk) with the copy to [tbi@dst.dk](mailto:tbi@dst.dk)

*5 size classes are defined according to both EU standards and in compliance with Armenian political needs.*

---

Employees	Armenian	EU standards
-----------	----------	--------------

needs		
10-29	10-29	10-49
30-49	30-49	
50-99	50-99	
100-249	100+	50-249
250+		250+

*Industry groups according to EU regulation*

Section C	Manufacturing
Section D	Electricity, gas and steam
Section E	Water supply, sewerage and waste management
Section F	Construction
Section G	Wholesale and retail trade; repair of motor vehicles and motorcycles
Section H	Transportation and storage
Section I	Accommodation and food service activities
Section J	Information and communication
Section L	Real estate activities
Division 69-74	Professional, scientific and technical activities
Section N	Administrative and support activities
Division 95.1	Repair of computers

**Descriptive table – data sets**

<b>Marz</b>	<b>Industry</b>	<b>Size class</b>	<b>Total employees</b>	<b>Number of enterprises</b>
Երևան	Nace C	10-29		
Երևան	Nace C	30-49		
Երևան	Nace C	50-99		
Երևան	Nace C	100-249		
Երևան	Nace C	250+		
Երևան	Nace D	10-29		
Երևան	Nace D	30-49		
Երևան	Nace D	50-99		
Երևան	Nace D	100-249		
Երևան	Nace D	250+		
Երևան	Nace E	...		
...	...	...		
Syunik	Nace X	10-29		
...	...	...		
Syunik	95.1	10-29		
Syunik	95.1	30-49		
Syunik	95.1	50-99		
Syunik	95.1	100-249		
Syunik	95.1	250+		

## **Annex 5 Statistics Denmark's analysis on Labour Survey data (sampling error)**

This paper gives a brief description of the quality of the sample for the Armenian Labour Survey. The analysis was done on data supplied for Armenian study visit to Statistics Denmark in late August. Only data concerning enterprises classified as small or micro was supplied, which means that bias assessments and sampling error estimates doesn't account for the fact, that all large enterprises probably was included in the sample as well.

### **Assessment of bias**

The population for the small and micro classified enterprises consists of 12051 enterprises; from which a stratified simple random sample of 4140 enterprises has been selected. The stratification is the combination of region, Nace code and size classification. Furthermore information on revenue and number of employees was supplied for every enterprise in population. To assess the bias we calculated sample estimates for both revenue and employees using the Horwitz-Thompson (H-T) estimator; which is given by

### **Formula**

These sample estimates was then compared to the true values, which were known since information on revenue and number of employees were supplied for every enterprises in the population. The deviations between the true values and sample estimates are of some concern. The sample estimates for revenue is approximately 27 pct. below the true value and the estimate for number of employees is approximately 40 pct. below the true value. Both of these results are quite extreme and very much exceed the random error stemming from the nature of sampling. This constitute a serious problem since the correlation coefficient between the study variable of interest (denoted  $h_{310_3}$  in the supplied data) and the number of employees is 0.61; This means that there's a very good chance that the sample estimates of the study variable of interest will be heavily biased.

### **Estimation of sampling error**

An estimate for sample mean of the variable  $h_{310_3}$  (in the formula above it would be  $\bar{y}_h$ ) and an estimate of the resulting standard error were also calculated. The formula for estimating the standard error of the mean is:

### **Formula**

Where  $f_h$  denotes the sampling fraction in stratum  $h$ , and  $s_h^2$  denotes the sample variance in stratum  $h$  given by:

### **Formula**

The estimated means and standard errors are given for the total and for each region below. All results

are for the study variable h\_310\_3;

### **Table**

The standard error for the total is quite acceptable. Especially when it's taken into consideration that it's only an estimate for the small and micro enterprises. The standard error for the total sample including all the large enterprises will probably be much lower. This is of course under the assumption that all larger enterprises is included in the sample and responds.

It's also a quite important point that these sample means differ from the sample means provided from

Armenia in a pivot table. A cautious guess would be that the provided sample means were un-weighted, and therefore didn't account for sample design.

### **Concluding remarks**

Even though the standard errors seem to indicate that relatively robust sample estimates can be calculated, then the fact that sample estimates of the control variables (revenue and number of employees) is so biased is such a strong indicator of an underlying problem with either the design or the

values of the control variables, that some sort of action to repair this has to be done.

The following should be examined with some scrutiny:

- If the design is in fact simple random sampling within strata

- The validity of the variables revenue and number of employees

- Alternative estimation techniques for instance calibration or combined ratio-estimation

The alternative estimation techniques should reduce the bias, but will probably increase the standard errors significantly.