



Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Nineteenth Meeting

Geneva, 4–6 October 2017

Item 4 of the provisional agenda

Evaluating the census and measuring data quality

**Quality Assessment of the Population and Housing Census
2014 in the Republic of Moldova**

Note by the National Bureau of Statistics of Moldova¹

Summary

In May 2014 the Republic of Moldova conducted the Population and Housing Census (PHC 2014). For the first time a Post Enumeration Survey (PES) was conducted and administrative data were used to evaluate the coverage and content errors of the PHC 2014. This paper shortly describes the methodology of conducting the PES, challenges and problems faced during the preparation and field operations, data linkage, use of electricity usage administrative data to complement PES results, some results regarding the coverage, conclusions and recommendations.

The overall PES and use of electricity data were carried out based on the technical and financial support of UNFPA, UN Population Fund, UNICEF and Swiss Agency for Development and Cooperation throughout the entire exercise.

¹ Prepared by Lilian Galer

I. Introduction

1. The PHC 2014, the second Population Census and the first Population and Housing Census, was conducted by NBS after the proclamation of the independence of the Republic of Moldova in 1991. According to the expert assessment recommendations of the 2004 Population Census and international recommendations of UN, for the first time, NBS decided to conduct a Post Enumeration Survey (PES) to assess coverage and quality of the PHC 2014. It was highly recommended to concentrate on the resident population coverage measures.

2. The final decision to conduct a PES was taken only 3 months before Census, without any test of methodology and questionnaire. Despite this, conducting of the PES was a very good opportunity for NBS to discover, beside coverage and content errors, other problems in preparing and carrying out of the Census. To complement the PES results, NBS also used electricity usage administrative data.

II. Quality evaluation of PHC 2014

A. PES as tool for coverage and content evaluation

3. The assessment of the quality of the PHC 2014 was done, primary, through the Post Enumeration Survey, which had as an objective the evaluation of both the coverage and the quality of collected data within census. This operation required a completely independent repeated data collection, from the same subpopulation, for a sample of Enumeration Areas (EAs), in different time occasions – during the PHC data collection and during PES data collection.

4. To ensure this independence:

- the PES was conducted after the census fieldwork was completed and Census material was collected in order to avoid contact between census and PES enumerators and possible transcription of data;
- the best census enumerators were selected as enumerators for the PES, but they were assigned to a different EA from the one they had enumerated at the census;
- Only the PES team knew in advance the distribution of the sampled EAs.

5. Entire PES work was led by the Statistical Methods Unit team together with the Social Statistics Department, which were not involved in the census data collection.

B. Sampling

6. The PES was based on a probabilistic stratified sample of 90 EAs, drawn from the list of all EAs used in census (approx. 11000 EAs). The population of interest was resident population living in residential dwelling on the territory of the Republic of Moldova at the time of the census, and did not include the individuals living in institutionalized dwellings and those living in the Transnistrian region of the Republic of Moldova, which is not under control of the Moldovan Government.

7. The population was stratified by geographic criteria in North, Center, South and Chisinau municipality and by residence. The sample size was determined mainly by financial and time constraints and the allocation was done proportionally with the population within each stratum. The sample was drawn using PPS sampling within each stratum, using the number of dwellings as size measure. An exhaustive re-enumeration of the dwellings/persons within selected EAs was done. The achieved sample size was of 21.7 thousands individual questionnaires.

C. Data collection

8. Data collection for PES was conducted in the period of 16-29 June 2014, 3 weeks after the census data collection finished.
9. The reference date for PES was the same as the one for Census- 12th May 2014 at 00.00 a.m.
10. For data collection PES questionnaires were used, which were a shorter version of the Census ones, with the addition of a short chapter for persons who left permanently/temporary the dwelling in the period among Census reference data and beginning of PES (out-movers) and persons who have moved in the dwelling temporary/permanently (in-movers).
11. To ensure the independence between PHC and PES in the fieldwork, the 90 enumerators for PES data collection were selected from other regions. They were grouped in nine mobile teams, each of them having a supervisor. A team was responsible for data collection in ten EAs. Each EA was divided in 10 parts on the map by supervisor.
12. The supervisors were experienced persons from the NBS Headquarter working with the household surveys. The supervisor coordinated the data collection of his team, defined the task for each day for the enumerators, checked daily the quality of collected questionnaires, helped the enumerators in some difficult situations and reported to the head office the daily progress in data collection.
13. After the collection, all collected data were checked once again in the NBS Headquarter by the PES team and stored within Statistical Methods Unit rooms, until the processing of the data.

D. Data processing

14. Unfortunately, due to various reasons, PES and Census data were stored for 23 months after data collection, without starting the data entry. Therefore, in April 2016 the data entry started for both PHC and PES. Double data entry for PES was done manually within an application written in CPro.

E. Data linkage

15. Data linkage is the most demanding and time consuming process from all PES activities. The missing of a unique identifier of the person in both PHC and PES, in the case of the Republic of Moldova Census, required to choose some variables that can show if there is the same individual or not. Hence, the variables used for linkage were: name, surname, address, date of birth, sex and household head.
16. Lacking of knowledge and experience in data linkage of the persons involved in PES resulted in a 7 months period of hard working without significant results in linkage. This happened because the first approach was to use the deterministic method of linkage. Because of big variation in the spelling of string variables (e.g. in PES - Cliciuc Ion, in PHC - Cliciuc Ivan; in PES – Sturza Adrian, in PHC Sturza Andrian, etc.) and unstandardized addresses (e.g. the name of the street - C MAKSA, C MARCSA, C MARX, CARL MARX, CARLA MARCS, KARL MARX, CARLO-MARCSA, etc.), there was a lot of time spent in standardization of the contents of the chosen variables. In spite of this huge work, the linkage rate was too low (around 65% of PES individuals).
17. In November 2016, UNFPA engaged an expert from Italian Statistical Institute which provided to NBS training and technical assistance in data linkage using RELAIS software. RELAIS offers the possibility to link data both using probabilistic and deterministic methods, thus the problem with different spelling was in most cases overcome. The steps taken for link the Census data with PES one was the following:

- a) Both files from PHC and PES was standardized in terms of variables names and special characters present in string variables (e.g. “;”, “””, “!””, “@”, etc.);
- b) Importing of the data in RELAIS software;
- c) Choosing of the blocking variable(s);
- d) Choosing between probabilistic vs deterministic model;
- e) Choosing of the matching variables;
- f) Choosing of the metrics used to measure the distance between values;
- g) Estimation of the model in case of probabilistic one;
- h) Reduction of the results to 1:1 linkage;
- i) Analyzing of the results by choosing the threshold (a probability) which indicate there is a link or not.

18. Choosing of the blocking variable, offer the possibility to restrict the space in which software will look for matches. From this point of view the linkage was done at different levels:

- at the level of EA;
- at the level of section (the second level in PHC data collection hierarchy)
- at the level of circumscripție (the third level)
- at the level of locality (the lowest level of administrative division in the Republic of Moldova)
- at level of raion (the first level of administrative organization).

19. After doing linkage using probabilistic model, there was used deterministic method (at the level of the country) and finally a visual (manual) check of the remaining non-linked cases. The results of linkage process are shown in table 1:

Table 1. Linkage of persons enumerated in PES and PHC 2014

Linkage at the level of:	Linked cases at the level of:			% of total linked cases		
	Chisinau	Rest of the country	Total	Chisinau	Rest of the country	Total
EA	2390	12449	14839	81.4%	79.9%	80.1%
Section	101	1035	1136	3.4%	6.6%	6.1%
Circumscripție	372	1500	1872	12.7%	9.6%	10.1%
Locality	20	364	384	0.7%	2.3%	2.1%
Raion		172	172	0.0%	1.1%	0.9%
Country	19	27	46	0.6%	0.2%	0.2%
Manually	34	32	66	1.2%	0.2%	0.4%
Total linked	2936	15579	18515	100.0%	100.0%	100.0%
Total PES	4660	17052	21712	63.0%	91.4%	85.3%

20. Therefore, during 4 months, NBS managed to link over 85% of the individuals from PES with PHC, using probabilistic (Fellegi-Sunter model) and deterministic methods.

21. After finishing the data linkage procedure, each individual was assigned a code describing the final status of the record – linked, unlinked and out of scope. For the estimation of the over-coverage rate, we used an alternative administrative data source – the electricity consumption per dwelling from the electricity provider. This data source has the limitation that can be linked with census data only based on addresses and name of one family member. Taking into consideration the very short period we had for linkage procedure and the limitation of the data source, there was taken decision to estimate the over-coverage only for Chisinau city.

22. The linkage of electricity data with PHC there was used also RELAIS, but the possibility to link data was restricted by available information in the electricity database. Hence, the linkage was possible only using the address and the name. The main problems arose with the address which are written in different ways in the both datafile (*see table 2*):

Table 2. The address spelling in PHC and electricity database

PHC			Electricity provider		
strada	nr	ap	strada	nr	ap
NICOLAE MILESCU SPATARU	23	39	Milescu-Spataru Nicolae str.	23	BLC39
NICOLAE SPATARU MILESCU	23	40	Milescu-Spataru Nicolae str.	23	BLC40
NICOLAE MILESCU SPATARU	23	41	Milescu-Spataru Nicolae str.	23	BLC41
NICOLAE MILESCU SPATARU	23	42	Milescu-Spataru Nicolae str.	23	BLC42
MILESCU SPATARU	23	43	Milescu-Spataru Nicolae str.	23	BLC43

23. Another problem was that as blocking variable we could use only administrative sectors of Chisinau city, and in the electricity data this one was not always correct. Due to these problems, in first attempts there was possible to merge automatically around 30.0% (76.6 thousands) of all addresses collected in PHC for Chisinau. After the standardization of the addresses in PHC data, the results were 58.4% (149.2 thousands) of all PHC addresses.

F. Estimation methodology

24. Estimation of the population was done using Dual System Estimation methodology, described in the UN PES Manual (2010). There were undertaken the following steps:

- assessment of the under-coverage (under-coverage rate);
- assessment of the over-coverage (over-coverage rate);
- estimation of the true population;
- estimation of the coverage indicators (census omission, erroneous inclusion and net coverage counts and rates).

25. The under-coverage rate and true population was estimated based on results of linked data from PES and Census data.

26. The over-coverage for Chisinau, as was mentioned above, was estimated based on the linked Census and electricity data for Chisinau. Based on the analysis of the linked data, we filled out the *table 1*:

Table 1. Contingency table of the counts in the Census and electricity data

		Electricity consumption		Total
		Present	Absent	
Census	Present	A	B	A+B
	Absent	C	D	C+D
Total		A+C	B+D	A+B+C+D

27. The population of interest in this case consisted only of persons enumerated in Census as residents, who assumingly used electricity. The subpopulation of linked dwellings was divided in two groups – with and without electricity consumption. We defined as over-covered (erroneous enumerated dwellings) all dwellings that were enumerated in Census, but during the year preceding Census did not have electricity consumption of at least 30 kW/month for a period of at least 7 months (B). The number of over-covered persons was estimated by summing up all persons enumerated in the linked dwellings with no consumption. The formula for estimation of the over-coverage rate was:

$$r_o = \frac{B}{A + B}$$

28. Of course, this approach underestimates the over-coverage, because there are taken into consideration only dwelling/households with all persons absent.

29. In absence of available information regarding over-coverage in other regions, we assumed that over-coverage in the country was at the same level as in Chisinau city, so the estimated over-coverage rate for Chisinau was used also for the rest of the country.

III. Results

30. The main results of the PES are, in terms of coverage:

- a) The PHC 2014 in the Republic of Moldova covered only 88.2% of total residents of the country, i.e., 11.8% of residents have not been covered, which represents an amount of 302.3 thousand residents, with the biggest under-coverage rate in Chisinau city (43.6%);
- b) At the same time PHC 2014 counted around 2.7% of the non-residents, who lived abroad for more than 12 months;
- c) The net under-coverage rate at the national level was 9.3%:
 - a huge difference in net under-coverage rate was registered between Chisinau city (41.2%) and the rest of the country (1.9%);
 - by residence, the net under-coverage rate varies between 0.6% in rural area and 22.4% in urban area;
 - there was registered a slight variation of the net under-coverage rate by age group from 6.1% for 65 year and more age group to 10.9% for 15-24 years age group. This situation is explained, from one side, by a younger population living in urban area and especially in Chisinau, which was highly under-covered, and, from the other side, by high mobility of this age group;
 - by sex there are no significant differences in net under-coverage;
- d) The number of estimated residents, at national level, according to the PHC 2014 was 2562.0 thousand persons, and according to the PES there was 2824.5 thousand persons living in Republic of Moldova on the reference day of 12 of May 2014, 83.7% of the net under-counts being from Chisinau city.

31. The big under-coverage in terms of persons in Chisinau city imposed NBS top-management to take the decision of correcting the population counts for Chisinau city from 312.9 to 532.5 thou persons.

IV. Problems and challenges of the PES

32. During PES implementation there was different problems and challenges NBS faced:
- a) There was a very limited time in order to properly design and conduct the PES – only about two months;
 - b) Because this was first PES, no any previous experience of the staff in conducting such kind of surveys;
 - c) Very limited staff resources – a total of 7 persons from NBS staff were involved in PES preparation and conducting, beside their main tasks. There was no staff fully dedicated to PES;
 - d) Unknown budget at the moment of decision taking to conduct the PES;
 - e) The preparation of the sample frame was done manually and took several days because there was not an unique database of census EAs. There were discovered some errors in assignment of territorial identifiers, which conducted to wrong stratification;
 - f) There was no time and possibility to conduct a pilot survey, before starting the field implementation;
 - g) The missing of addresses in small rural communities (less than 5000 inhabitants) and low quality maps raised some problems in EAs identification in the field – in case of 14 EAs there were used lists instead of maps, but with strong guidelines to include all new dwellings found in the area;
 - h) The use of mobile teams ensured the complete independence of the data collection, but increased the probability of overlapping between the parts the EAs were divided. Also, there were cases when respondents refused to cooperate because the enumerator was not from their locality;
 - i) There were some attempts by local authorities to obstruct de data collection for the PES;
 - j) Lack of the pilot imposed 3 or 4 changes of the daily monitoring system in the first 3 days;
 - k) Due to the delay in data processing, there was not possible to conduct the reconciliation visits in order to clarify the status of the problematic linkages;
 - l) Manual data entry can be an additional source of errors in final data set;
 - m) A lot of spelling errors in string variables resulted in a waste of time and resources, without good results, using the deterministic method of linkage at the beginning;
 - n) The probabilistic linkage was done de facto by only two persons during 4 months, working 12 hours a day.

V. Problems in PHC 2014 identified through PES

33. Beside evaluation of the Census quality in terms of coverage and content errors, the PES also identified some problems in conducting PHC:
- a) The delineation of EAs was done mainly based on the list of dwellings and not on the map, which raised some problems in EAs identification, especially in rural area where there are not addresses;
 - b) Where maps where used for the delineation of EAs, sometime they were of poor quality;
 - c) There was not created a database with all EAs. The info was kept in separate Excel files, which conducted to duplication of EAs identifiers. The identifiers for some EAs were changed during data collection;

- d) There were situations when the limits of the EAs were changed during the data collection;
- e) The manual data entry did not allow to have a good control over the ID of the persons through the roster and individual questionnaires.

VI. Conclusions and recommendation

34. Based on the experience and the results of the PES conducted in connection with the 2014 PHC, the following conclusions and recommendations can be considered:

- a) Preparation for the next Population and Housing Census and PES should start earlier with detailed and adequate planning of human, technical and financial resources;
- b) For PES should be allocated human resources dedicated only to it;
- c) The preparation of the fieldwork (mapping, address system, methodology for EAs delineation, GIS) should start immediately, because it will be a time consuming activity;
- d) All work on EAs delineation should be done using GIS technologies and most updated available information from Cartographic Agency at the address point level;
- e) It is highly recommended to include the reconciliation visits as the phase of the PES in the initial planning;
- f) For the assessment of over-coverage it can be also used other alternative sources (e.g. Labor Force Survey);
- g) A system of daily monitoring for both PHC and PES should be developed

VII. References

POPOVSKA-TOSHEVA Ljubinka, Technical report on “Technical Assistance to the National Bureau of Statistics to design, manage and implement the Post-Enumeration Survey of the Population and Housing Census 2014 in Republic of Moldova”, UNFPA project “Enhanced capacity of national institutions to undertake the Population and Housing Census”Skopje 2016

TUOTO Tiziana, Mission Report on “Estimates for population size and quality indicators for Moldovan Population and Housing Census 2014”, UNFPA project “Enhanced capacity of national institutions to undertake the Population and Housing Census”, Rome 2017

UN (2010), Post Enumeration Surveys. Operational guidelines. Technical Report, Published by the United Nations Secretariat, Department of Economic and Social Affairs, Statistics Division: New York, April 2010.