

**Economic and Social Council**Distr.: General  
12 July 2017Original: English  
English and Russian only**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****Census quality control (performance, lessons learned and directions for the next PES)****Note by the Croatian Bureau of Statistics <sup>1</sup>***Summary*

Croatian Bureau of Statistics simultaneously carried out the coverage and quality control of the 2011 Census, providing that the quality control covered 10% households out of the number selected in a sample. Soon after the completion of the 2011 Census in the field, the Post Enumeration Survey (PES) was carried out in order to estimate the coverage of the Census. In assessment of census data coverage dual system of estimation has been used. By using certain topics, Central Bureau of Statistics (CBS) wanted to estimate the accuracy of responses in regard to particular questions in questionnaires as well as to control the methodological correctness of responses according to given topics.

The PES made possible to assess coverage errors of the 2011 Census as well as the quality of responses that were included in the response quality control. There are certain directions derived from the experiences CBS got in this PES.

Applied methods for evaluating census error in 2021 will differ from those applied in 2011 and 2001.

<sup>1</sup> Prepared by Damir Plesac



## Contents

	<i>Paragraphs</i>	<i>Page</i>
I. Introduction .....	1–2	3
II. Post Enumeration Survey (PES) in general and the aim of the PES in Croatia.....	3–5	3
III. Sample size.....	6	3
IV. Review of main results .....	7–17	4
A. Persons.....	8–13	4
B. Households .....	14–15	4
C. Dwellings.....	16–17	4
V. Errors found in the Census coverage and their sources .....	18	5
VI. Human resources and their challenges in the PES .....	19–22	6
VII. Comment on the PES questionnaires .....	23–33	6
VIII. Performance of the PES versus Census .....	34–35	9
IX. Applied methods for evaluation of Census Error in 2021 .....	36–37	9
X. Conclusions (Summary of lessons learned from errors, conclusion and directions for the next PES).....	38–39	10
References .....		11

## **I. Introduction**

1. Considering that census data and their various implementation are of great strategic importance, the Croatian Bureau of Statistics makes great efforts in order to ensure a full coverage of the Census of Population, Households and Dwellings. Nevertheless, no matter how great these efforts may be, no census so far managed to cover all enumeration units, persons in the first place.

2. In this document, the greatest attention is given to coverage of the 2011 Census. This document also discusses the general features and objectives of the PES in Croatia, sample size, the main findings, human resources and their challenges in the PES, comment on the PES questionnaires, summary of lessons learned from errors, conclusion and directions for the next PES.

## **II. Post Enumeration Survey (PES) in general and the aim of the PES in Croatia**

3. The aim of the coverage control of enumeration units was to check the coverage of all enumeration units and their correct allocation, that is, to see if some of them were accidentally dropped out.

4. The aim of the response quality control was to evaluate the reliability of responses to particular topics from the questionnaires. The response quality control was also used for the control of the methodological correctness of responses according to the required topics, that is, for checking up of the consistency in the implementation of the prescribed instructions, definitions and classifications as determined by the Census methodology.

5. CBS aimed at starting the PES as soon as possible after the Census, for the shorter the time between the execution of the Census and PES is the better PES quality is achieved. Therefore, training of the enumerators and controllers was carried out on 29–30 April 2011 and enumeration in the field started on 2 May 2011, which is only four days after the completion of the Census in the field, and lasted until 10 May, or nine days.

## **III. Sample size**

6. Out of slightly more than 26 000 enumeration areas in the 2011 Census, about 22 700 or 87% of areas i.e. 99% of households, were part of the basic set from which a sample was selected to participate in the Post//enumeration survey. Enumeration areas with less than 30 households in urban areas were excluded from the basic group and the same was done with enumeration areas with less than 15 households in the non-urban stratum, due to the large share (35%) of enumeration areas with less than 30 households. Due to the fact that the sample size depended on timely and financial limits as well, the coverage control was carried out on the sample containing 300 enumeration areas, or 1.4% of households in the Republic of Croatia. Out of 300 enumeration areas, 154 were included in the rural stratum, 146 in the urban one (out of which 56 in the City of Zagreb).

## IV. Review of main results

7. The main conclusions were:

### A. Persons

8. In the Republic of Croatia 98.1% of population have been covered by 2011 Census of Population, Households and Dwelling, i.e. 1.9% of population of Croatia have not been covered by 2011 Census of Population, Households and Dwelling, which amounts around 81 000 persons.

9. Net undercoverage of population was 1.75% or 73 000.

10. Net undercoverage of male was 1.82% or 37 000 and net undercoverage of female was 1.68% or 36 000.

11. If observing population coverage by stratum, than:

- Net undercoverage of population in urban areas was 1.99% or slightly more than 46 000;
- Net undercoverage of population in rural areas was 1.44% or slightly more than 27 000.

12. While the 2011 Census determined that the Republic of Croatia has about 4 285 of inhabitants, the results of the Post Enumeration Survey suggest that at the reference moment on 31st of March 2011 there were about 4 358 persons with usual residence in the Republic of Croatia.

13. There is a small variation among sexes. The highest percentage of population not covered by Census was among younger population, aged 15–24 and 25–39. This is understandable knowing the fact that young people are more involved in migration flows (internal and international) and more often change residence. It mostly regard students, people aged 19–27, who are absent from their place of permanent residence almost entire year, and young people who try to become independent from their parents or search their own dwelling to live in and form the family. The highest coverage was among the youngest age-group 0–14.

### B. Households

14. In the Republic of Croatia 98.39% of households have been covered by 2011 Census, i.e. 1.61% of households of Croatia have not been covered by 2011 Census.

15. If observing households coverage by stratum, than:

- Undercoverage of population in urban areas was 1.98%,
- Undercoverage of population in rural areas was 1.09%.

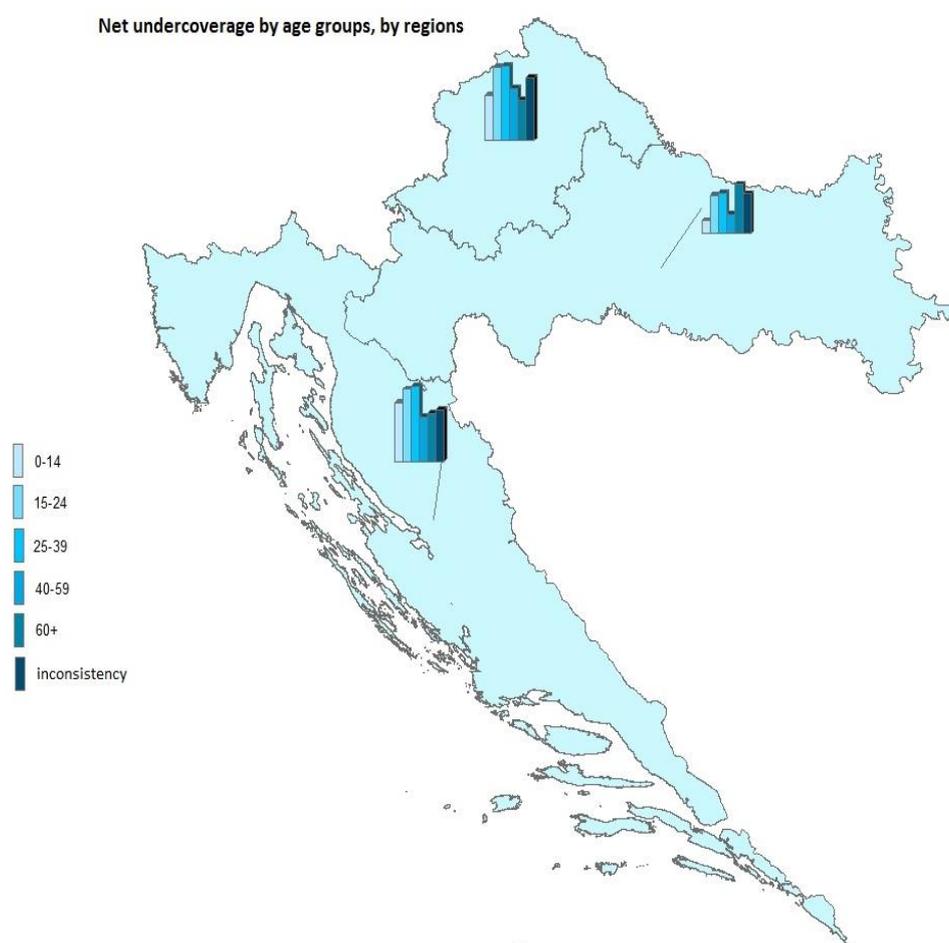
### C. Dwellings

16. In the Republic of Croatia 96.44% of dwellings have been covered by 2011 Census, i.e. 3.56% of dwellings have not been covered by 2011 Census.

17. If observing households coverage by stratum, than

- Undercoverage of dwellings in urban areas was 3.70%;
- Undercoverage of dwellings in rural areas was 3.42%.

Figure 1  
**Net undercoverage by age groups, by regions**



## V. Errors found in the Census coverage and their sources

18. In such a large, comprehensive and complex action as the Census, oversights in the coverage of persons and dwellings cannot be avoided. In Croatia PES revealed various reasons of omitting the persons and dwellings in Census:

- (a) Prior to the very enumeration, an enumerator failed to complete the visit of his enumeration area and to check its limits according to the documentation he got and therefore he failed to enumerate all the persons and dwellings located within the limits of his enumeration area (mainly in Zagreb area);
- (b) Temporarily empty dwellings were characterized as abandoned;
- (c) There were persons who either deliberately escaped the Census or reluctantly responded to questions (out of their fear that these information may be used against their best interests);
- (d) There were persons did not want to open the door to unknown visitors;

- (e) There were several households living in one dwelling, so it happened that a particular person in a household or the whole household were left out;
- (f) Persons were temporarily absent (e.g., at work, at school);
- (g) Persons with no place of usual residence and who, due to various reasons, did not come to enumeration centers to be enumerated.

## **VI. Human resources and their challenges in the PES**

19. One of the unavoidable links in the PES chain are persons that participate in the project in the field during data collection and in the office during data processing. In the 2011 PES, persons, who took part in data collection during the Census, were also engaged in the PES. In order to ensure the greatest independency possible, there have been taken certain steps, as follows:

- It was forbidden that enumerators participating in the PES were selected for the territory they were selected in the Census;
- Data collected in the particular area during the Census were not available to PES enumerators;
- The PES could be started only after collecting and storing all census questionnaires to specially controlled premises;
- The sample of enumeration areas selected into the PES was sent to the field only after it was sure that persons engaged in the PES in a particular area could not reach data on that area.

20. Although these measures helped in protecting the Census data very well and the data independence in the field achieved, according to CBS opinion there are still several considerable reasons to avoid engaging in the PES the same persons that participated in the Census:

- After four weeks of hard work in the Census, in some regions a great number of high quality and well trained controllers did not want to work in the PES and had to be persuaded to get involved into the PES, while in some other regions controllers competed among themselves to get a job on the PES;
- Some persons who, after the Census, continued to work in the PES lost their willingness and necessary concentration. Their work came down to wishing to finish the work as soon as possible (not taking into account its quality).

21. In order to achieve more quality in the PES than in the Census, person working in it should be highly motivated and extremely well trained, which is possible only in the case they have enough time for training and if they are more than sufficiently awarded for their work. It must be said that in this case the compensation for the PES work was only "sufficient" (at the same level), as compared to the compensation for the Census.

22. Concerning data processing, positive side was in precisely defining the staff and time necessary for all segments of the PES work, in order to avoid the situation in which a part of work must be delayed because other works, that should already be done, are late.

## **VII. Comment on the PES questionnaires**

23. In preparing the PES questionnaires, CBS leaned on the contents of the Census questionnaires, providing that they had to be as simple as possible. The Census and the PES

questionnaires were processed by using the scanning technique (optical recognition of characters).

24. The document outlines some ideas for improving the quality of future data collection in the Post-enumeration Survey (PES). It is observed that much of the literature about a PES does not deeply address PES questionnaire design. It is generally the case that while the importance of the PES questionnaire is acknowledged, specific details on the design are absent. For example, the United Nations report (2010) states that: “The PES questionnaire plays a central role in the survey process in which information is transferred from the respondents to the survey analysts”.

25. Here we concentrate on potential nonsampling errors resulting from item missing data. Note that as with unit refusals item refusals can also lead to item nonresponse bias. For each item in a survey it is possible to estimate the item nonresponse bias through the following equation:

$$INRB(y_i) = \left(\frac{NR_i}{N_i}\right)(y_{Ri} - y_{NRi})$$

where:

- *INRB(y<sub>i</sub>) is the item nonresponse bias for item y<sub>i</sub> in the PES;*
- *NR<sub>i</sub> is the number of missing questions for item y<sub>i</sub>;*
- *N<sub>i</sub> is the sample size;*
- *(NR<sub>i</sub>/N<sub>i</sub>) is the item nonresponse rate or missing data rate;*
- *y<sub>Ri</sub> is the estimate of the total based on those that respond to the specific question;*
- *y<sub>NRi</sub> is the estimate of the total based on those that did not respond to the question y<sub>i</sub>.*

*Note that if the item nonresponse rate is zero, that is, if  $NR_i/N_i = 0$  then the nonresponse bias is zero.*

*It is also zero if  $y_{Ri} = y_{NRi}$ .*

*It is important to note that this equation can also be written specifically for missing data based on those that say they do not know an answer to a question and also for those that refuse to answer a question*

26. As a first step in understanding the potential magnitude of this source of error it is recommended that the Croatian Bureaus of Statistics (CBS) quantify the missing data rates for each question in the 2021 PES survey.

27. Question 14 of the PES questionnaire (KP1) was “Who provided the data to the enumerator?”, where the possible responses were:

- (a) The person him/herself;
- (b) Household member;
- (c) Another person.

28. In KP2 form the enumerator is required to enter information about who provided the information about the dwelling (since this question was not numbered in KP2, it will be called question 14a in the sequel). That is, whether the information was provided by household member or other person. The exact question and response categories were:

Table 1

**Question 14a**

<i>Who provided the data to the enumerator?</i>	<i>1. Household member</i>
	<i>2. Another person</i>

29. This information applies to specific questions on the dwelling unit including whether the housing unit is private property or has some other status, the size of the unit, the units usage, the surface area, number of rooms in the dwelling, etc.

30. It is interesting to note that question 14 and 14a have different response categories, given that the household respondents (KP2 form) gave the person who was, at the time, the head of the household. These questions (14 and 14a) are particularly important to understanding the quality of the PES. As a first step the tabulation of the responses to question 14 and 14a should be examined by PES post strata. For question 14 the preferred respondents are either the person him or herself or another household member and for question 14a the preferred respondent is a household member. Examining this data by post-strata can provide valuable insight into the 2021 PES, especially for enumerator training if some post-strata have large numbers of respondents that are not the preferred respondent in the PES. Then missing data rates by post strata and type of respondent can identify important areas that might require improved questionnaire design, more detailed enumerator training, etc.

31. Upon review of the 2011 PES questionnaire it is observed that enumerators cannot code a reason why a question was not answered – was it because the respondent did not know the answer or because the respondent refused to answer? These two different reasons for item nonresponse could lead to different types of nonresponse bias. Thus it is important to know why item missing data occurs, especially because in 2021 Census and 2021 PES much higher number of refusal in answers is expected than in 2011, where the number of refusal was negligible.

32. The following suggests how “don’t know” (DK) and refusal (R) might be accounted for in the design of future PES questionnaires. This example is extracted from KP1 form (All new design suggestions are in bold font).

Table 2  
**Extract from the KP1 form**

1. Sex	1. Male 2. Female
2. Date of birth	<b>Enumerator: If Don't Know or Refused circle appropriate code, 98 or 99</b> Day / month / year _ / _ / _ _ _ _ <b>98 DK 99 R</b>
3. Identification number	PIN <b>98 DK 99 R</b> ID <b>98 DK 99 R</b>
4. Marital status	Single, married, widowed, divorced <b>98 DK 99 R</b>
5. Number of children born alive	1,2,3, four and more (state the number), none <b>98 DK 99 R</b>
6. Education attainment (not to be answered by pre-school children)	1. No schooling 2. Uncompleted basic school Highest grade the person has completed: 1,2,3,4,5,6,7 <b>98 DK 99 R</b> Please see the other codes in Form 1, Question 22
7. Did the person carry out any activity (at least 1 hour in the period 25–31 March for profit, in cash or in kind?)	Yes No, but has a job to return to No <b>98 DK</b> <b>99 R</b>

8. Did the person seek a job during March?	Yes
	No
	<b>98 DK</b>
	<b>99 R</b>
9. Is the person available to start working in the next two weeks?	Yes
	No
	<b>98 DK</b>
	<b>99 R</b>
10. Has the person ever worked before?	Yes
	No
	<b>98 DK</b>
	<b>99 R</b>
11. Status in employment	Please see codes in Form 1, Question 29
12. Occupation	PES answer _____
	<b>98 DK</b>
	<b>99 R</b>
13. Activity status in the period 25–31 March	Please see codes in Form 1, Question 32
	<b>98 DK</b>
	<b>99 R</b>
14. Who provided the data to the enumerator?	1. The person him/herself
	2. Household member
	3. Another person

33. For questions 14 and 14a the working hypothesis is that the lowest missing data rates occur when the person (him or her) is the respondent, followed by a household member and that the highest missing data rates occur when “another person” responds to the PES.

## VIII. Performance of the PES versus Census

34. The question why the PES performed more or less at the same level of quality as the Census did, when the original intention was that it should perform better. Some factors are that field staff have been tired after their previous work on the census, and that some people have gone home specially for the census and left again shortly afterwards, therefore being more likely to be missed in the PES. There is also the factor that some people refused to take part in the PES, they were mostly in the Zagreb city area.

35. In general, the PES appears to have succeeded in its most important task: that is providing the data for good estimates of census coverage. This has been possible because it was carried out independently of the census, using field staff who were new to their areas and who had no knowledge of the information, which had been gathered in the census. It is this independence, which justifies the use of dual system estimation.

## IX. Applied methods for evaluation of Census Error in 2021

36. Up to now, in Croatia, PES was the only source of measuring the quality of census, in particular the coverage and the quality of data content of the population and housing

census. Since Croatia has a wide range of administrative data sources associated with various government activities, such as: Birth and death register, register of insured persons and register of beneficiaries of pension insurance rights from the Croatian Institute for Pension Insurance; registry of Persons with Disability of the Croatian National Institute of Public Health, evidence of permanent and temporary residence of the Ministry of Interior, register of insured persons of the Croatian Institute for Health Insurance, State Geodetic Directorate database, PIN system within the Ministry of Administration, e – registra, a centralized system of the Ministry of Science, Education and Sport, etc. all this administrative data sources shall be used for statistical purposes and among them for measuring the quality of 2021 census.

37. Nevertheless, the main challenge encountered when using administrative data for statistical purposes will be the specificity of the administrative purpose and a lack of reference to the statistically defined ‘resident population’.

## **X. Conclusions**

### **(Summary of lessons learned from errors, conclusion and directions for the next PES)**

38. In general, it could be said that the PES reached the required aims. Although all persons engaged in the Census 2011 did their work thoroughly, they could not remove all the errors, which cannot be avoided in such a large statistical action as it is the Census. The PES made it possible for CBS to assess errors in the coverage of the Census 2011 as well as the quality of responses that were included in the response quality control. Therefore, it offers a picture of the value and accuracy of Data, which proved to be of immense importance in many cases. The quality control of responses to certain questions, that is, the defining of the number of errors, takes into account the assumption that, along with the census results, which are expected to have errors, there is another set of responses that are considered true. Measuring the volume of errors comes to the comparison of particular results in these two sets. The true set of responses should be obtained by the PES. There lies its basic problem. Can be sure that responses obtained by the PES are true and those in the Census false? Especially if we bear in mind that CBS does not have a chance to check the differences between the PES and Census 2011 responses. Therefore, CBS does not dispose of a complete proof that the PES responses are true. So, it is assumed that the PES and the Census responses are approximately equally true and that is why dual assessment system has been chosen.

39. As a census without any error does not exist, there is still the need for the PES as a part of census programmes. In many countries it is considered the most reliable method for the assessment of census results. In order to make the PES more successful in the future, there are certain directions derived from the experiences CBS got in this PES:

- (a) Defining deadlines for all stages of the PES and stick to them;
- (b) Choosing of effective and realistic matching rules;
- (c) Pointing out, as much as possible, the idea of independence of the PES implementation in relation to the Census;
- (d) Standardising of working concepts and definitions of the PES and the Census;
- (e) Making sure that every information included in the matching between the PES and the Census is useful;
- (f) Inclusion of well trained and quality persons for the work in the field;

- (g) Well instructing of the main staff, which takes part in the selection of samples, on the methods of sample survey;
- (h) Get necessary resources for the PES within resources defined for the Census implementation;
- (i) Keeping the PES as less complicated as possible and stick to reachable aims;
- (j) Checking, in the field, all aberrations, preparing of proof tests for such cases.

## **References**

- B. Mahon, (2001) Training course on post enumeration surveys for the Republic of Croatia.
- D. Plesac, (2015) Coverage and content control of the 2011 Census of Population, households and dwellings, study and analysis.
-