

Distr.: General
11 September 2017

English

United Nations Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Nineteenth Meeting

Geneva, Switzerland

4–6 October 2017

Item 4 of the provisional agenda

Evaluating the census and measuring data quality

Lessons learned from household data quality assessment in the 2011 Czech Census

Note by the Czech Statistical Office¹

Abstract

The paper presents the main results of the household census data quality assessment and focuses on methodological and technological innovations and aspects in the last population and housing census 2011. Specifically, factors such as transition to the usual residence approach, new data collection methodology, and fieldwork quality are assessed. The results point out the importance of fieldwork methodology as well as the input data quality which are the key factors of household structure data quality. Also possible over/underestimation has been calculated. Lastly, the paper indicates lessons learned from the assessment (such as encourage response via internet) as well as introduces the plans and main innovations in 2021 census.

¹ Prepared by Pavlína Habartová, Census Methodology, Analyses and Dissemination unit

I. Background

1. Compared to previous population censuses, the 2011 Population census brought several significant changes, such as e-forms, mail-back collection data option, outsourced fieldwork, usual place of residence approach, population register usage as an alternative data source etc. However, the traditional method of enumeration still prevailed.
2. While some of these changes and innovative approaches have significantly improved the data quality, some of them have limited it. Generally speaking, e-forms have been considered a breakthrough, which has definitely enhanced the data quality. On the other hand, mail-back option reduced considerably the data quality and increased the data processing demands.
3. The forms were separated for individuals, dwellings and buildings with the intention of allowing people to choose a collection method independently on other household members. However, since the address register at the level of dwelling does not exist in the Czech Republic, this friendly approach of forms collection implied higher demands on data processing, namely the creation of a person-household-dwelling-building links.
4. All these factors have affected in a certain way household structure data. The aim of this article is to assess the extent of possible effect of methodological and technological innovations and to propose recommendations for 2021 Census in order to minimize negative impacts on data quality.

II. Household data processing in the 2011 Population Census

5. In accordance to European regulations for Census round 2010 the place of usual residence as a territorial identification has been implemented. Equally, dwelling occupation status as well as household concept has been also based on usual residence.
6. The key step preceding the type of household derivation was linking persons with dwellings. On the basis of place of usual residence every person was linked with a building and dwelling, and then several corrections had to be accomplished before type of household derivation.
7. For the first time in the Czech census history, the household type derivation was developed to be fully algorithmized and programmed without any need of trained operators. The algorithm consisted of 5 steps, starting from determining a number of family nuclei, and then following by identification of type of family nuclei, number of children and deriving their dependency, and finally household and family coding. Besides, the algorithm determined the head of household and household status of every individual. Another important part of household derivation was also creating household datasets.

III. Household data quality assessment

8. Population and Housing Census 2011 used many innovative approaches, both technical methodological ones, often enforced by international requirements. While some had a clear positive effect on census results, some caused

complications or not provide data, which give an accurate picture of reality. In the field of household surveys, there are several key points that could influence household structure data. In this case, it has been multi-channel data collection methodology, a new usual residence approach, and automated data processing.

IV. Methodological aspects

9. For the first time since 1970, Census 2011 did not allow monitor time series of three types of households (dwelling, housekeeping, census). The housekeeping household instead of so called census household became the smallest processing unit. As the number of census and housekeeping households has been converging and as early as 1991, 98% of households were consisted of single census household, information loss was not significant.

10. Another important methodological breakthrough was the usual residence concept. The main idea of the concept was to capture the factual number of the population and households, not only in the territorial detail but also in terms of more reliable household structure. More than 90% of the population was enumerated at the place of usual residence. Almost 89% of population had the place of usual residence equal to the place of permanent residence. The largest differences were found at the age of 24–30 years.

11. Due to a strong correlation with age, the housing trend outside of the place of permanent residence also reflects household status structure (see figure 1). While the highest match of both addresses has been observed for persons in the position of spouse and grandparent with children, the most significant variance was found (with the exception of people not in private households) for persons in consensual unions (29%) and multi-member non-family households (27%). For these types of households, the effect of usual residence approach has been considered the most significant. In the case of consensual unions in the positive sense, in the case of multiple households in the negative one.

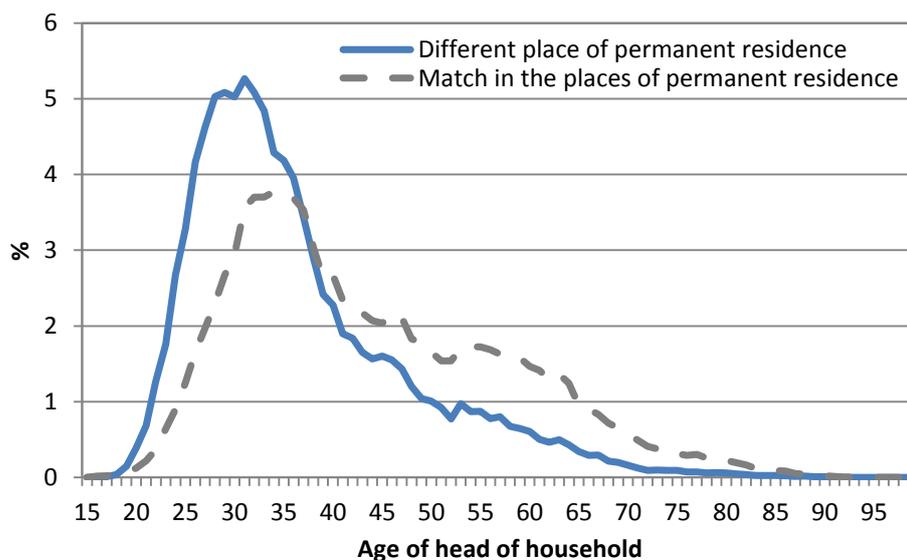
12. While in previous censuses the number of consensual unions was underestimated by the condition of a common place of permanent residence, the 2011 Population Census allowed for the first time in the modern census history to capture even couples with different places of permanent residence. In 2011, 40% of partners in consensual unions had a different permanent residence. In other words, if the permanent residence concept had been used in 2011, as was the case of previous censuses, only 60% of the consensual unions would have been possible to capture.

Figure 1
Persons by household status and the match in usual and permanent place of residence

Household status	Non-match in usual and permanent place of residence	Match in usual and permanent place of residence
Husband / wife in a married couple	5.7	94.3
Partner in an opposite-sex consensual union	29.3	70.7
Partner in a same-sex registered union	18.2	81.8
Partner in a same-sex consensual union	22.0	78.0
Lone parent	10.3	89.7
Grandparent living with grandchildren	4.1	95.9
Child	7.9	92.1
Grandchild living with grandchildren/grandchild	12.2	87.8
Other person not member of the family nuclei, but in a nuclear family household	16.0	84.0
Person in the household consisting of two and more family nuclei's	7.7	92.3
Person living alone	16.3	83.7
Person in a multi-person non-family household	27.4	72.6
Person in a institutional family household	61.8	38.2
Person in a institutional non-family household	40.2	59.8
Primary homeless person	79.2	20.8
Total	10.8	89.2

Source: Population and housing census 2011

Figure 2
Age distribution of consensual unions by match in the places of permanent residence of both partners



Source: Population and housing census 2011

V. Data collection and technological aspects

13. The new way of fieldwork data collection (using a contractor) and the automatic data processing more likely than the household methodology affected the results on household structure. The household derivation algorithm was based on the assumption of good input data quality, primarily correctly filled in the ID's. However, these assumptions have not been fulfilled, mainly due to the high proportion of forms sent back by mail (60%). Since the identification of the forms was not filled in correctly, severe problems in the linkage of all three entities (individuals-dwellings-buildings) were faced.

14. During the data processing it was found that the dwelling number (which was the most important part of every ID) was completely missing in 3% of dwelling forms and 11% of individual forms. The poor quality of the enumerators' fieldwork in filling in the dwelling number (fictitious dwelling ID) has been also illustrated by the fact, that in 41% of the cases the dwelling number had to be corrected during data processing. Subsequently, since the alternative procedures had to be used for individual allocation and entities linkage, households based on de facto marital status more likely than de iure status became difficult to capture.

15. Another negative impact had the way of capturing the input data concerning the list of persons living in the same household. This list of persons was scanned but not verified, which to a certain extent limited the following process of household processing.

VI. Using registers and automated allocation to households

16. Another important innovation of the last census, which could affect the household structure, was the usage of population register for supplementing some missing attributes as well as for completing whole missing data records. In total, more than 3% of personal records have been added from the population register (figure 3). Mostly, these persons were allocated as a member of non-family multi-person household (26%), as an individual in one-person household (19%) and as a child (16%). In other words, nearly one fifth of people in multi-person non-family households were made up of persons who were supplemented from the population register. The most frequent reason was the fact, that no direct relatives at the place of permanent residence have been found.

17. All the key processes preceding the type of household derivation such as linking entities, and person allocation were successfully automated. However, it has led to a false increase in the number of non-family households for such cases when no other persons with relatives at the same dwelling have been found. Data quality analysis showed significant divergence between family and non-family households in terms of their common place of usual residence. While in almost 90% of family households all members have been enumerated at the place of usual residence, only one quarter of non-family households consists of members with the equal place of enumeration and place of usual residence; which also confirms the overestimation of the non-family households.

Figure 3
Individuals based on population register records only (not enumerated in the field) by household status

Household status	Persons based on register record only (not enumerated)	
	in % of total number of persons with the household status	in % of total number of persons based on register record only
Husband / wife in a married couple	0.9	9.5
Partner in an opposite-sex consensual union	0.3	0.4
Partner in a same-sex registered union	2.8	0.0
Partner in a same-sex consensual union	1.4	0.0
Lone parent	3.2	5.1
Grandparent living with grandchildren	0.3	0.0
Child	2.1	16.3
Grandchild living with grandchildren/grandchild	1.1	0.1
Other person not member of the family nuclei, but in a nuclear family household	8.8	10.9
Person in the household consisting of two and more family nuclei's	1.1	1.1
Person living alone	4.7	18.8
Person in a multi-person non-family household	18.1	26.0
Person in a institutional family household	2.5	0.1
Person in a institutional non-family household	22.3	11.7
Primary homeless person	0.0	0.0
Total	3.3	100.0

Source: Population and housing census 2011

VII. Comparison with other data sources

18. Apart from census data, also other data sources on households such as EU-SILC and LFS can be used for data quality analysis. However, it is important to keep in mind different methodology, which by nature leads to inconsistencies in household structure comparison. While sample surveys show a higher proportion of one-couple families than the number based on the census data, the share of one person households is estimated at a significantly lower level than in the census data; even at a lower level than was estimated by the census data in 1991 and also 9% below the EU average.

19. On the other hand, the comparison with the sample surveys also pointed out the overestimation of the number of multi-person households in the census data. As the analysis showed, the overestimation can be considered at the expense of one-couple family household, more specifically cohabiting couples.

Figure 4

Household distribution by data source, 2011

Type of household	Census	EU-SILC	LFS	Census – EU average (median)
One-couple family household	47.9	63.9	59.8	52.3
Lone parent household	13.0	11.4	9.3	10.1
Multi-person non-family household	4.9	1.0	2.7	2.7
One-person household	32.5	23.7	28.2	32.6

Source: Population and housing census 2011, EU-SILC 2011, LFS 2011, Census hub

VIII. Discussion, lessons learned and challenges for 2021 Census

20. The quality data assessment showed that the place of usual residence approach had the most significant positive impact on the household structure in terms of methodological innovations. On the other hand, the major negative effect was observed for the fieldwork quality (especially the mail-back option), which caused considerable problems in data processing. Also using population register in terms of supplementing the individual records to the data set did not have only positive impact on the data quality, since it resulted in considerable share of not stated responses and necessary imputations. The automated procedure of assigning people to dwellings and households itself, logically limited the quality of the outputs, however, the algorithm was exhausting enough to be considered very successful. Finally, the programmed algorithm for type of household derivation derived only the type of household according to the algorithm and had no direct effect on the resulting household distribution.

21. The analysis showed that the census data most likely overestimated the number of multi-person non-family households by almost 50%. On the contrary, some underestimation can be assumed for the number of one-couple families with younger children, including consensual unions, despite the usual residence approach. Conversely, the number of one-person households can be considered very close to the reality. The results pointed out that published data on household structure is reliable and consistent with the historical trends, as well as comparable to the international demographic trends and households formation.

22. Drawn on the basis of the result of the quality assessment findings and lessons learned, some recommendations towards the improvement of 2021 population and housing census data quality have been:

- a) Emphasis on the fieldwork quality,
- b) No mail-back option for paper forms collection,
- c) Increase in the online participation (encourage response via internet)
- d) No separate forms for household and household members (i.e. common household questionnaire for whole household and its members),

- e) Validation of scanned names and surnames of household members in the household matrix,
- f) Using more registers and administrative data sources for usual place of residence derivation,
- g) Improvement of household matrix questions towards the intelligibility,
- h) Increase the response rate.

23. Since the preparation of the 2021 census is already in progress, most of these recommendations have been already implemented into the baseline features of the next census. The main issue is both the fieldwork quality and the online census, which are limited by the absence of the address register to the dwelling level. Therefore, the return to common census form for the entire dwelling household, which was abandoned in the 2001 census due to scanning and data protection issues, is again considered.

24. Another important challenge is the assessment of household composition questions and the survey method (item g)), in order to increase questions clarity as well as the associated higher quality of the input data. However, both the demand on space on the form and the reduction of response burden are the key factors. Also the question of possible corrections which could limit unreal number of some households will be considered.

IX. Conclusions

25. The 2011 Census has brought some innovations, which could have both negatively and positively affected the data quality. The analysis showed that the usual residence approach has positively affected the data quality. Most data variations were observed already in input data; therefore, the quality could not dramatically improve during automated processing, which followed. However, the results are still comparable to historical and international trends and overall deviations are minimal.

26. Some of the recommendations which have been drawn on the basis of the analysis results have been already implemented into the baseline features of the next census. The most important are the emphasis on the fieldwork quality, no mail-back option for paper forms collection and the validation of scanned names and surnames of household members in the household matrix. Nevertheless, all of them can be solved by sufficient response rate via internet which is the main aim of the next census.

27. In 2016, CZSO began the preparation of the 2021 Census Act. In particular, according to the draft Act the basic principles of the census 2021 are: using of administrative data sources in the greatest possible extent, reducing the amount of paper forms as well as the response burden, timeliness, the data protection etc.

References

Habartová, P.: Families and households in census data with the focus on the methodological aspects of the data. 2016. PhD Thesis.

Population and housing census data, 2011, Czech Republic

EU-SILC data, 2011, Czech Republic

LFS data, 2011, Czech Republic

2011 Census hub, available here:

<<https://ec.europa.eu/CensusHub2/query.do?step=selectHyperCube&qhc=false>>