

Using Mobile Positioning for Improving the Quality of Register Data

Kaja Sõstra¹ and Kristi Lehto²

¹Statistics Estonia, e-mail: kaja.sostra@stat.ee

²Statistics Estonia, e-mail: kristi.lehto@stat.ee

Abstract

The next population and housing census in Estonia in 2021 is planned to be register-based. One of the biggest challenges on the way to register based census in Estonia is the difference between registered and actual places of residence. By law, everybody is obliged to ensure that their correct usual address is entered in the Population Register, but there are different reasons why people don't do that. The difference between registered and actual places of residence affects the regional population statistics and all household and family characteristics.

To solve this problem Statistics Estonia (SE) carried out a pilot project for testing possibility to use mobile positioning data. The anchor points of the first and second place of residence were estimated based on the mobile positioning data. Anchor points and other auxiliary information was used to build a model for selecting the most probable place of residence from the set of addresses.

Keywords: register based census, mobile positioning, anchor point model

1 Introduction

Countries who have conducted register based population and housing census (PHC) are using information of place of usual residence from Population Register. Estonia is planning to conduct the next PHC based on registers. Estonian Population Register (PR) contains information of a place of usual residence. The problem is, that people don't live in their place of residence registered in PR in Estonia. Surveys performed by Statistics Estonia show that 20–25% of all permanent residents of Estonia have not registered their actual place of residence in PR and live at a different address. This problem has historical roots. At the beginning of Estonia reindpendence in 1990s the registration of a place of residence in PR was voluntary. Now it is obligatory more than ten years but lots of people don't consider registration necessary (Äär, 2017). Today, there are several new reasons and factors causing false registration of place of residence in PR: school and kindergarten places, free public transport, several social benefits, wish to support different local government etc.

The difference between registered and actual places of residence affects the breakdown

of the regional population statistics and all household and family characteristics. (Tiit, Visk, Levenko, 2018).

2 Pilot study

2.1 Participation in pilot study

Statistics Estonia started pilot project for testing possibility to use mobile positioning data for register based census. The aim of pilot study was to test the feasibility of specifying actual place of residence with the help of mobile positioning data. Study was organised in cooperation with University of Tartu and Positium LBS. The pilot study consists of following steps:

- Statistics Estonia asked volunteers to participate in the pilot
- Set of potential addresses was created for each participant based on registers
- The home anchor points were estimated based on the mobile positioning data
- Anchor points and other auxiliary information was used to build a model for selecting the most probable place of residence from the set of addresses.

All participants of the study filled in digitally signed written consent to use their mobile positioning data. Together with consent, the following information was collected: participant's name and identification code, mobile number and mobile operator, the actual place of residence. Identification code was used for linking data from registers, mobile number and mobile operator were needed for obtaining mobile positioning data from operator. Participant's actual place of residence was collected to compare positioning data and register data with actual address.

All together 310 persons participated in the study, four participants gave two phone numbers. The phone numbers divided between three mobile operators: Telia (158) Elisa (103), Tele 2 (53). Mobile positioning data was received from Telia and Elisa. Participants with mobile numbers of operator Tele 2 were not included into next steps of the study.

2.2 Linking of addresses from registers

Set of potential addresses was created for each participant based on Population Register and Land Register. The following addresses were linked to the participant's data:

- Place of residence of participant from Population Register (PR)
- Addresses of participant's real estates from Land Register (LR)
- Place of residence of participant's spouse and children from PR
- Addresses of real estates of participant's spouse and children from LR
- Addresses of other relatives from PR and LR

Addresses from registers were compared with participant's actual place of residence. The actual place of residence coincided with PR address for 67% participants. For participants with different actual place of residence and PR address other addresses were compared with place of residence. Most likely these persons live on the address of own real estate registered in Land Register (26%), on the address of spouse's real estate (19%) or spouse's address in Population Register (14%). For about half participants living elsewhere than PR address it was possible to find their actual address from registers.

2.3 Calculation of anchor points

Agreements with mobile operators, mobile phone positioning data extraction and anchor points calculations were made by company Positium LBS. Mobile phone positioning data was received from two mobile operators (Telia and Elisa) for 261 phone numbers and 257 persons. Anchor points were calculated for 243 phone numbers (93%) and 240 persons. The mobile phone usage activity of some persons was too low for calculating anchor points.

Anchor points were calculated according to methodology developed by University of Tartu (Ahas et al, 2010). Anchor points are locations which person visits regularly. Calculation of anchor points consists of eight steps. The two regular cells that had the highest number of days with calls are selected for the calculation of home and work-time anchor points. Home anchor point has average starting time of daily calls after 17:00 and standard deviation of beginning time of calls greater than 0.175. Working time anchor point has average starting time of daily calls before 17:00 and standard deviation less than 0.175. If the home and work-time anchor points are located in the same network cell and cannot be separately identified then anchor point is determined as work-home anchor point.

For estimating the quality of anchor points, the home and work-home anchor points were compared with the actual place of residence of participant. Anchor point and actual place of residents were in the same county – 97%, in the same municipality – 87%, in the same settlement – 67%. The coordinates of the actual place of residence were inside of anchor point polygon for 82% of participants. Comparison shows that coincidence is high. Most of cases the distance is shorter than 10 km. There are only few cases with large discrepancy between the anchor point and the place of residence.

The accuracy of anchor points is at the level of the service area of a mobile antenna (a network cell). The diameter of the polygon of anchor point is negatively correlated with population density. There are more mobile antennae in the densely populated areas. Diameter of polygon of anchor point is about 1-1,5 km in urban areas and 10-15 km in rural areas.

3 Validation of the quality of place of residence in registers

For comparison and making decisions the polygon of anchor point and coordinates of register address were used. If polygon of home, second home or work-home anchor points included coordinates of register address then it was decided that register address is actual

address. If coordinates of register address were outside of all polygons then it was decided that this address is not actual address of participant. This decisions were compared with information about actual place of residence.

Results of validation of participants PR addresses are shown in table 1.

Table 1: Results of validation of PR addresses

| | | Decision made using mobile positioning data (home, work-home and second home anchor point) | | |
|---|--------------------|---|--|--------------|
| | | PR address inside anchor polygon | PR address outside anchor polygon | Total |
| Coincidence of actual place of residence with PR address | PR = Actual | 146 | 15 | 161 |
| | PR ≠ Actual | 16 | 63 | 79 |
| | Total | 162 | 78 | 240 |

Correct decision is made based on mobile positioning data for $146 + 63 = 210$ persons (87%). Other register addresses described in chapter 2.2 were compared similarly and decision was made if other register address is actual address. The most probable place of residence was selected from the set of potential addresses.

4 Conclusions

One of the biggest obstacle on the way to register based census in Estonia is the problem of difference between registered and actual places of residence.

Feasibility study showed that mobile positioning data could be one solution for solving this problem. It is possible to use mobile positioning data for validation of the quality of registered addresses. One prerequisite for using mobile positioning data for register based census is solving legislation issues.

References

- Ahas, R., Silm, S., Järv, O., Saluveer, E., Tiru, M. (2010), Using mobile positioning data to model locations meaningful to users of mobile phone. *Journal of Urban Technology*, 17:1, 3-27
- Sõstra, K., Lehto, K. Geospatial mobile data to increase the quality of usual place of residence. *Paper presented in European Conference on Quality of Official Statistics Q2018*.
- Tiit, E.-M., Visk, H., Levenko, V. (2018), Partnership index. *Quarterly Bulletin of Statistics Estonia*. 1/2018, pp.29 - 43
- Äär, H. (2017), Coincidence of actual place of residence with Population Register records. *Quarterly Bulletin of Statistics Estonia*. 1/2017, pp. 80–83.