Evaluation of the quality of population register data in Slovenia by conducting a targeted survey on plausible emigration

Note by the Statistical Office of the Republic of Slovenia

Summary

The main aim of the survey conducted by the Statistical Office between March and June 2016 was to evaluate above all the rate of over-registration in the Central Population Register (CPR), which is the basic administrative source on the population for the register-based census in Slovenia, which was for the second time conducted in 2015. Some methods used to measure over-registration in Slovenia are also presented (residence status of selected respondents and their household members from sample surveys, register-based labour force status derived annually). The main part of the paper presents the methodology for determining the sample frame for conducting the special targeted survey on possible non-registered emigration based on register-based labour force status and both methods used for collecting data (postal method, paper assisted personal interviewing - PAPI). The high non-response rate (71%) in the postal survey was expected. The main outcome from the postal and PAPI surveys is confirmation of relevant quality of CPR data for population counts from international perspective as the overestimation of population at the aggregate level is less than 0.5% (around 10,000 persons). Future steps on possible changes in the production of regular quarterly demography statistics and census counts are still under consideration and will be decided after a complete evaluation of the survey.

1 Prepared by Danilo Dolenc.
I. Introduction

1. The Central Population Register (hereinafter the CPR), which was established prior to the 1981 Census by the Statistical Office, has a long tradition in Slovenia from administrative and statistical point of view. The CPR was transferred for keeping and maintaining to the Ministry of the Interior in 1998 as Article 59 of the National Statistics Act adopted in 1995 determined three years of transition period.

2. Administratively speaking the registers in general are intellectual wealth of a nation, but the CPR is even more than that – it covers the complete life cycle from birth to death or in other words the history of every individual tracing his/her life-time events. The basic administrative quality principles for maintaining the population register are (MNZ, page 15):

   - Complete coverage (all events are included irrespective of population status or citizenship)
   - Officiality of the data source (recorded by authorized persons)
   - Individual responsibility for correctness of data (together with auditing in databases)
   - Stability of data sources (maternity hospitals, other health institutions, courts, territorial register, local registry offices)
   - Uniqueness of coverage using a key identifier (EMŠO = PIN) (no duplication of records)

3. Data from the CPR are widely used in the public administration with direct or indirect connection to the CPR data in more than 20 agencies and governmental bodies. Permanent, long-term, widespread uses of register data and systematic update of errors detected are the most important factors of quality of population register data.

4. CPR data were for the first time used for the population counts as of 1 July 1985 following the administrative definition of population. From 1995 to 2007 a modified definition of population was in force but still mostly based on administrative concepts. The real statistical definition of usual population was applied in 2008 affected by two facts:

   - Establishment of e-CPR, which means that the Statistical Office receives all data needed for the statistical process on population (more than 150 variables) on a regular (quarterly) basis.

5. For producing demographic statistics the so-called Demographic Database was established. Population counts based on statistical definition converting administrative concepts into statistical data are produced quarterly and are at the same time the backbone of the register-based census which was carried out twice (in 2011 and 2015). Around 98% of records from the basic e-CPR table as of 1 January 2015 (the main criterion is a valid registered residence in Slovenia) belong to population, 1.4% are records for persons living abroad but still having a valid registered address in Slovenia and 0.6% are records for persons living or intending to live in Slovenia less than a year.

II. Methods for measuring over-registration

6. We know from the theory of registers that the most critical part of data sources used for updating the population register are individuals reporting (or not) the changes of their
residence in the country and above all emigration abroad. Over-registration is therefore a common problem of all countries with register-based statistics. The following methods for measuring over-registration of the population register that could be used in Slovenia are described briefly in the following paragraphs.

7. The sample frame for regular sample statistical surveys is usual population based on statistical definition of population from the Demographic Database as of 1 January, 1 April, 1 July or 1 October available 4 months after the reference date. Due to the time span between sample extraction and field operation (at least 5 months) the sample is updated before the field operation started with information on dead and emigrated persons (excluded from the sample) and with internal changes of addresses (not excluded from the sample).

8. The following information on the selected respondent from the sample is collected in every social sample survey and results are presented in Table 1:

- Does the selected respondent live at the sampled address?
- If not, the reason why he or she no longer lives there is collected with the categories:
  - Has moved
  - Dead
  - Unknown
- If the answer is “Has moved”, the additional information is collected with the categories:
  - To another private household in Slovenia
  - To an institutional household in Slovenia
  - To abroad
  - No information on present address or address is unknown.

Table 1
Selected respondent person according to the residence status (share in % of sample), different sample surveys, Slovenia

<table>
<thead>
<tr>
<th>Survey</th>
<th>Total</th>
<th>Died</th>
<th>Unknown</th>
<th>Living elsewhere</th>
<th>Abroad</th>
<th>No answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBS 2012</td>
<td>5.3</td>
<td>0.2</td>
<td>0.5</td>
<td>4.6</td>
<td>3.6</td>
<td>1.0</td>
</tr>
<tr>
<td>HBS 2015</td>
<td>9.4</td>
<td>0.2</td>
<td>1.0</td>
<td>8.2</td>
<td>5.5</td>
<td>1.8</td>
</tr>
<tr>
<td>LFS 2014</td>
<td>8.8</td>
<td>0.1</td>
<td>1.4</td>
<td>7.3</td>
<td>5.8</td>
<td>1.5</td>
</tr>
<tr>
<td>SILC 2014</td>
<td>6.2</td>
<td>0.7</td>
<td>0.5</td>
<td>5.0</td>
<td>2.6</td>
<td>1.6</td>
</tr>
<tr>
<td>SILC 2015</td>
<td>8.4</td>
<td>0.9</td>
<td>0.5</td>
<td>7.0</td>
<td>4.2</td>
<td>1.9</td>
</tr>
<tr>
<td>ICT-HH 2014</td>
<td>7.9</td>
<td>0.3</td>
<td>0.8</td>
<td>6.8</td>
<td>4.9</td>
<td>1.3</td>
</tr>
</tbody>
</table>

HBS - Household Budget Survey
LFS - Labour Force Survey
SILC - Statistics on Income and Living Conditions
ICT-HH - Survey on Information and Communication Technology usage in households
Source: SURS, internal documentation.

9. The higher share of dead persons in SILC is the consequence of the panel survey. But as we also update the sample monthly during the field operation in most cases we can inform the interviewers in advance about deaths, emigrations and internal migrations. From Table 1 we can conclude that internal differences between de facto and registered residence
are the most relevant for the quality of territorial distribution of population data (3% to 6%). According to data from sample surveys the over-registration in the CPR and consequently the population is overestimated by 1% to 2% (persons living abroad).

10. The opposite approach (presented in Table 2) is proper just for internal distribution of population and could be traced in household surveys only. The point here is the share of persons not having administrative residence in the interviewed household but being the household member according to the survey. As expected the share is almost half lower (2.1% to 3.2%) than from the previous method presented in Table 1.

Table 2
Interviewed persons in household surveys according to the administrative registration, sample surveys, Slovenia

<table>
<thead>
<tr>
<th>Type of administrative residence</th>
<th>SILC 2014</th>
<th>SILC 2015</th>
<th>HBS 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Share (%)</td>
<td>Number</td>
</tr>
<tr>
<td>Household members - total</td>
<td>28,176</td>
<td>100</td>
<td>26,571</td>
</tr>
<tr>
<td>Registered residence in the household</td>
<td>27,287</td>
<td>96.8</td>
<td>25,773</td>
</tr>
<tr>
<td>Residence registered elsewhere</td>
<td>889</td>
<td>3.2</td>
<td>798</td>
</tr>
</tbody>
</table>

Source: SURS, internal documentation

11. The labour force status derived from 9 administrative and statistical sources in register-based population = census statistics was described in detail in the paper Deriving Labour Force Characteristics from Multiple Sources in the Register-Based Census of Slovenia, presented at the Seventeenth Meeting of the Group of Experts on Population and Housing Censuses in Geneva in 2015. The imputation rate for labour force status (no data on individual person in any of 9 sources, only data from the CPR are available, so there is a high probability that those persons live abroad) could be used for measuring the overestimation of population. Besides that, the number of records with imputed value taken over from the previous year (which is decreasing year by year) is a sign of improved quality of the CPR. The most problematic are data on foreign citizens (they represented 5% of total population as of 1 January 2015), above all those with permanent residence in Slovenia. The quality of data on administrative survivors (persons registered as still alive in the CPR but already dead in fact) is improving and also the number of persons aged 95+ is insignificant.

Table 2
Labour force status – imputation rates, Demographic Database, Slovenia, 1 January 2011-2015

<table>
<thead>
<tr>
<th>1 January</th>
<th>Imputed records from previous year</th>
<th>Imputation rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>2011</td>
<td>-</td>
<td>1.50</td>
</tr>
<tr>
<td>2012</td>
<td>23,919</td>
<td>1.52</td>
</tr>
<tr>
<td>2013</td>
<td>23,509</td>
<td>1.50</td>
</tr>
<tr>
<td>2014</td>
<td>20,546</td>
<td>1.29</td>
</tr>
<tr>
<td>2015</td>
<td>17,276</td>
<td>1.17</td>
</tr>
</tbody>
</table>
12. For comparison, the difference in counts from both sources in 2002 (CPR = 1,994,861; Census 2002 = 1,964,036) was 1.6%, which is even higher than any imputation rates for labour force status.

II. Targeted survey on plausible emigration from Slovenia – the sample

13. The survey on potential non-registered emigration from Slovenia was conducted in the scope of the Eurostat grant “Usual Residence Population Definition: Feasibility Studies”. The criteria for selecting the focused sample were also prepared on the basis of methods described before. The following criteria were taken into account using hierarchical order as some records could be selected by more than one criterion:

(a) No data on labour force status in any source for three consecutive years;
(b) Foreign citizens with permanent residence in Slovenia without data on labour force status in any source in the previous year;
(c) Citizens of the Republic of Slovenia with temporary residence in Slovenia only and without data on labour force status in any source in the previous year.

14. On the other hand, we also detected a population group which is marked in administrative data (CPR) as non-residential group but data on their labour force status in Slovenia could be found in at least one out of nine sources. Persons belonging to this group are citizens of Slovenia mostly having permanent residence in Slovenia and at the same time they have registered temporary residence abroad and according to the statistical definition of population they are excluded from the population of Slovenia (almost 30,000 persons in total, the vast majority without information on labour force status). On the basis of data on labour force status we presumed that some of them live in Slovenia and should be treated as residents but they forgot to register their return from abroad. So this group represents the only under-estimation in population counts not taking into account possible illegal (undocumented) immigrants who are not detected even in any sample survey.

15. 15,490 persons aged 15 years or more according to the statistical definition of population (reference date 1 January 2015) correspond to the first three criteria described above and another 2,703 persons were sampled from the fourth group which is excluded from the population stock. The total sample therefore consists of 18,193 persons. As the field survey started in the middle of March 2016, the following procedures were applied to prepare the final sample for the field inquiry using the latest available administrative data as of 1 March 2016:

• In case of at least two household members or persons living in the same conventional dwelling only one was selected by the seniority criteria;
• Persons who died or emigrated abroad in the meantime were excluded;
• Foreigners whose temporary residence permit had expired were excluded.
### III. Targeted survey on plausible emigration from Slovenia – content and methods

16. A relatively short questionnaire was prepared (see Annex 1, Survey questionnaire Residing abroad, 2016) with the following topics to be answered by sampled respondents:

- Information on person who filled in the questionnaire (we were aware in advance that the selected person is absent from the country in many cases, so other persons will possibly reply);
- International migration experience (country of residence, year of immigration/emigration, reason for migration);
- For persons living abroad the intention to return to Slovenia;
- For persons living in Slovenia the intention to emigrate from Slovenia in the following year (country, reason).

17. Two methods for collecting field data were applied. Firstly, the postal method was chosen as we anticipated that many letters will be returned and not delivered to the addressee (unknown person at the address). The prepaid envelope for returning the filled in questionnaire was enclosed. Letters were sent out in the middle of March 2016 with a two-month deadline (15 May 2016).

18. As we expected a high non-response rate, we conducted a field inquiry (face to face interview) using PAPI for practical reasons (short questionnaire, relatively small sample of 1,915 persons, easy data entry). Only 7 interviewers did the job in one month (June 2016) as the sample was determined by the territory first and all persons living there were included in the survey. The decision for such a method based also on the rationality to decrease expenses (at the most travel costs) and time needed to carry out the task. The areas for the survey were chosen on the basis of the return rate from the postal method, the number of persons included in the final sample and also on the basis of knowledge of possible high undocumented emigration:

- Selected quarters of Ljubljana and Maribor (areas with the highest number of sampled persons) – 3 interviewers in Ljubljana (885 respondents) and one in Maribor (317 respondents);
IV. Targeted survey on plausible emigration from Slovenia – preliminary results

19. As we are still receiving some responses, the data presented in the paper are preliminary (but very close to the final ones) and also data editing was not completed at the time of finalizing the paper, so only the most important variables for evaluating the quality of population register data are analysed here.

20. The response rates were very low but better than expected. The non-response rate (no information on the sampled person) was 71%, irrespective of the criteria used for sampling. The undelivered letters returned directly by the Post Office to the NSI (1,632) were branded with their official postmark and labelled as addressee is dead (21), moved from this address but no information where he/she is living now (445) or is unknown at the address (1,166). We have to take into consideration that the Post Office in Slovenia runs its own system for mail delivery and that its information on the actual situation in the field is regularly updated and trustworthy. The highest response rate was 25% for the persons temporary living abroad as many respondents from this population group actually live in Slovenia.

Table 5 Response rates by inclusion criteria, postal survey, Slovenia

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Nr. of respondents</th>
<th>Letters returned by Post Office</th>
<th>Filled-in questionnaires</th>
<th>Non-response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>14,219</td>
<td>11.5</td>
<td>17.5</td>
<td>71.0</td>
</tr>
<tr>
<td>Over-registration – total</td>
<td>11,678</td>
<td>13.6</td>
<td>16.0</td>
<td>70.4</td>
</tr>
<tr>
<td>No data on LFS three consecutive years</td>
<td>10,359</td>
<td>13.3</td>
<td>15.9</td>
<td>70.8</td>
</tr>
<tr>
<td>No data on LFS previous year (foreign citizens)</td>
<td>1,041</td>
<td>19.0</td>
<td>14.2</td>
<td>66.8</td>
</tr>
<tr>
<td>No data on LFS previous year (citizens of the RS)</td>
<td>278</td>
<td>5.0</td>
<td>24.1</td>
<td>70.9</td>
</tr>
<tr>
<td>Living temporary abroad but LFS data available</td>
<td>2,541</td>
<td>1.6</td>
<td>24.6</td>
<td>73.8</td>
</tr>
</tbody>
</table>

LFS – labour force status

21. The most interesting or target outcome of the survey is division of responses and determination of resident status of respondents into the following categories:

• Respondent actually resides abroad - overestimation in population statistics;
• Respondent has died but is not registered in the CPR (administrative survivors) - overestimation in population statistics;
• Respondent actually lives in Slovenia – underestimation in the case of persons having permanent residence in Slovenia but excluded from population counts as at the same time they have registered temporary residence abroad;
22. Not taking into account not identifiable resident status from the postal and field survey (in most cases the interviewer had not succeeded to make any contact with the selected respondent in the cities of Ljubljana and Maribor), the results from the postal survey and the field survey are comparable to a large extent. If the respondents belonging to the population stock are the subject, the ratio between over-estimation and correctness is 72.5% to 27.5%. If we generalize that, almost three quarters of persons not having information on labour force status in any source being part of population counts are not living in Slovenia. But if we take into consideration all records from the survey, just over half of responses (56.1%) were categorized as over-estimation, almost one in five (18.8%) as under-estimation and exactly one in four (25.1%) as correct (respondent actually lives in Slovenia and is included in population counts or respondent actually lives abroad and is not included in the population of Slovenia). The ratio between over-estimated population and under-estimated population in the survey was 3 to 1, so the over-estimation in total survey is in fact exactly half of absolute error (2,160+719 = 2,879; 2,160-719 = 1,441). If we take into account also the correct records according to the statistical population definition (958) and we assume that the resident status of the total sample is distributed as found in the field survey, the final over-estimation of the population is around half of the value of the indicator calculated from the total sample of the survey and the sample frame (population aged 15 years or more), which is 1.03%.

Table 6
Resident status by type of survey and type of respondent, Slovenia

<table>
<thead>
<tr>
<th>Resident status</th>
<th>Total</th>
<th>Postal survey</th>
<th>Field survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>All persons</td>
<td>All persons</td>
<td>Population</td>
<td>All persons</td>
</tr>
<tr>
<td>Total</td>
<td>6,037</td>
<td>5,009</td>
<td>4,122</td>
</tr>
<tr>
<td>Over-estimation</td>
<td>2,160</td>
<td>2,160</td>
<td>1,380</td>
</tr>
<tr>
<td>Living abroad</td>
<td>2,042</td>
<td>2,042</td>
<td>1,286</td>
</tr>
<tr>
<td>Already died</td>
<td>118</td>
<td>118</td>
<td>94</td>
</tr>
<tr>
<td>Under-estimation</td>
<td>719</td>
<td>-</td>
<td>522</td>
</tr>
<tr>
<td>Correctness</td>
<td>958</td>
<td>813</td>
<td>608</td>
</tr>
<tr>
<td>Not identifiable</td>
<td>2,200</td>
<td>2,036</td>
<td>1,612</td>
</tr>
</tbody>
</table>

V. Evaluation of the quality of population register data and lessons learned

23. The number of so-called administrative survivors (persons who haven't been removed from the CPR but should be due to death) is negligible (98), of which 25 are aged 85+. According to the paper (Oblak Flander, 2009), the number of centenarians in population statistics in 2007 in Slovenia was overestimated by 15% to 20%. As of 1 January 2015 the administrative survivor rate calculated from the survey results was maximum 8.5% (if all 20 out of 236 centenarians included in the survey are dead as we received only 4 responses reported death out of 20 selected persons aged 100+).
24. The most problematic administrative data from the over-estimation point of view are data on foreigners with permanent residence in Slovenia as many of them do not deregister themselves before departure abroad. The main reason is the status of having permanent residence in an EU Member State (as most of them are citizens of non-EU countries from the territory of former Yugoslavia), which gives many rights to the holder. Besides that, administrative authorities pay much less attention to the quality of data on foreigners compared to the data on citizens of the Republic of Slovenia.

25. A small population group of citizens of the Republic of Slovenia with temporary residence in Slovenia only without labour force status data (1,300 persons) should be investigated more deeply as the survey results prove that most of those persons are living abroad.

26. A positive effect of the survey was also detected in the administrative source. After sending letters to respondents (middle of March 2016), by 1 August 2016 (latest available CPR data) in only four months almost 10% of person from the total sample deregistered their residence in Slovenia. 83% of persons included in the survey did not respond to the NSI but just to the administrative authority.

27. Statistically speaking, the data on total population produced according to the statistical definition of population based on administrative CPR data are of excellent quality. The overestimation of population at the aggregate level is estimated to be less than 0.5%, which means around 10,000 persons as maximum. Even at the individual level the total error rate is something between 0.7% and 0.8%. Compared to the last field census data in 2002 (Dolenc, 2003), where almost 1% of the population was counted twice (over-coverage) and slightly less than 2% of the population was not counted at all even though they were living in Slovenia (under-coverage), the results of the survey are promising. We are convinced that any field sample survey (not to mention a census!) returns worse quality data than those from the Central Population Register in terms of coverage, relevance, accuracy, timeliness and coherence.

28. Of course, there is still room for improving the quality of CPR data mostly by the administrative authority. In July 2016 a new Residence Registration Act was adopted with some new solutions how to achieve the main intention of the law - proper registration of de facto residence of persons living in Slovenia. The results obtained from different methods presented in the paper proved that the quality of input CPR data is better and better every year. There is still a delay between de facto and registered emigration abroad but it is shorter than before. The close cooperation between the NSI and the administrative authority is an important key on that road.

29. Future steps on possible changes in the production of regular quarterly demography statistics and census counts are under consideration. We already excluded reported dead persons in the survey from population counts. Whether or not to exclude also the population with successive item non-response in all sources used for the production of labour force status will be decided after a complete evaluation of the survey.

References

Dolenc D. (2003), Register based 2002 census of population, households and housing in Slovenia and new solutions in data processing, WP.4, Joint ECE-Eurostat Work Session on Population and Housing Censuses, Ohrid

Dolenc D. (2015), Deriving labour force characteristics from multiple sources in the Register-based Census of Slovenia (Slovenia), WP.14, Joint UNECE-Eurostat Meeting of the Group of Experts on Population and Housing Censuses, Geneva

MNZ (2015), Centralni register prebivalstva Republike Slovenije: število posredovanih podatkov je doseglo 1.000.000.000, Ministrstvo za notranje zadeve, Ljubljana
Survey questionnaire
Residing abroad, 2016

Who is answering the questionnaire?

1. Addressee.
2. Other family or household member.
3. Other person.
4. The addressee died.
5. The addressee rejects participation.
6. The addressee doesn’t live at this address – lives elsewhere in Slovenia.
7. The addressee doesn’t live at this address – lives abroad.
8. The addressee is unknown at this address.
9. Other: ________________________________
    ________________________________
    ________________________________

End of questionnaire.
Thank you!

Residence **outside of Slovenia** is defined as residence **outside of Slovenia**, **without**:
- traveling, vacation or visits to relatives (regardless of duration);
- residence abroad due to work or study, if you return to Slovenia weekly.

Have you ever resided or are you currently residing outside of Slovenia?

1. Yes, but I am currently living in Slovenia.
   
   Enter the year of immigration/return to Slovenia: __________

2. Yes, I have been living outside of Slovenia for **at least 1 year**.
   
   Enter the year of leaving Slovenia: __________

3. Yes, I have been living outside of Slovenia for **less than 1 year**.

4. No, never. → continue with V7
In which foreign country did you last reside or are you still residing?

- [ ] Austria
- [ ] Bosnia and Herzegovina
- [ ] Croatia
- [ ] Italy
- [ ] Kosovo
- [ ] Macedonia
- [ ] Germany
- [ ] Serbia
- [ ] Switzerland
- [ ] Other:

Who is/was residing abroad with you?

- [ ] Nobody.
- [ ] Spouse or partner.
- [ ] Spouse or partner and children.
- [ ] Children.
- [ ] Parents.
- [ ] Other people.

What is/was the main reason for residing abroad?

- [ ] Employment.
- [ ] My employer sent me to work abroad.
- [ ] Study, work placement during study, internship.
- [ ] Family reasons (marriage, family reunification, etc.).
- [ ] I was born in a foreign country.
- [ ] Other:

Answer question V6 only if you are currently residing abroad, otherwise continue with question V7.

Do you intend to return to Slovenia?

- [ ] Yes, within one year.
- [ ] Yes, within 5 years.
- [ ] Yes, after more than 5 years.
- [ ] Yes, but I don't know when.
- [ ] I don't know yet.
- [ ] No.

Do you intend to leave Slovenia within the following year?

- [ ] Yes.
- [ ] No.
- [ ] I don't know.

Which country are you planning to move to?

- [ ] Austria
- [ ] Bosnia and Herzegovina
- [ ] Croatia
- [ ] Italy
- [ ] Kosovo
- [ ] Macedonia
- [ ] Germany
- [ ] Serbia
- [ ] Switzerland
- [ ] Other:

What is the main reason for thinking of moving abroad?

- [ ] Employment.
- [ ] My employer will (probably) send me to work abroad.
- [ ] Study, work placement during study, internship.
- [ ] Family reasons (marriage, family reunification, etc.).
- [ ] Other:

Thank you for completing the survey!