Overcoverage in the Total Population Register

Tor Bengtsson
Stina Åsling Rönning
Statistics Sweden
Overcoverage in the Total Population Register

Tor Bengtsson\textsuperscript{1}, Stina Åsling Rönning\textsuperscript{2}

The main quality risk for the Total Population Register (TPR) is the existence of overcoverage in the register. We can assume that persons who are living in Sweden leave their imprints in different administrative registers. For example, they can have income, study or find a new place to live. Persons who do not have imprints in the registers can then be assumed to have moved out of the country without reporting this. In order to find out the number of overcoverage, annual information from internal registers based on administrative registers from other government agencies have been used to identify those without imprints. When deciding whether a person belongs to the overcoverage group or not, a number of different indicators are used. The indicators are created by a set of rules that can be based on register data from a single year or from several years, and even from rules concerning what happens following years. Because certain rules indicate overcoverage better than others a weighted count is preferable.

The number and the characteristics of the overcoverage is presented, and since the overcoverage will have different impact on statistics based on a survey, where TPR is the frame, and statistics based on registers a discussion on these differences is held.

The Swedish Tax Agency has been able to use information about the overcoverage to make the control of the population registration more efficient and the results had an impact on the population statistics. The control led to a decrease in the number of the population in some municipalities that normally have population growth.

\textbf{Key words:} overcoverage, Total Population Register, register based statistics, non-response, estimation.

\section{Introduction}

The main quality risk for the Total Population Register is the existence of overcoverage in register. It happens when people registered in the population

\textsuperscript{1} Statistics Sweden, Senior Statistician, Örebro, Sweden, Email: tor.bengtsson@scb.se

\textsuperscript{2} Statistics Sweden, Senior Advisor, Örebro, Sweden, Email: stina.ronning@scb.se
register leave the country to settle permanently abroad without being registered as emigrants. Therefore the register includes more people than it is meant to cover, which normally has no impact on the quality of the statistics. But for certain domains and topics it may cause problems. This may for example lead to difficulties when studying the integration process for immigrants.

2. Background

For at least the last two decades attempts of various kind have been made in a number of studies to estimate how large errors in the registers are for immigrants living in Sweden. Common types of estimation have involved investigating the proportion of returned mail, the non-response rate in surveys, studies of people who lack a recorded income and comparing mortality levels for immigrants compared to those of people born in Sweden. Overcoverage in the TPR has an impact on the sample process, level of non-response and may cause unexpected problems in the process of estimation.

3. Overcoverage

3.1 Number of persons without imprints

The latest attempts to estimate the level of overcoverage have been based on the assumption that persons who are living in Sweden leave their imprints in different administrative registers. For example, they can have income, participate in studies or find a new place to live. Persons who do not have any imprints at all in the registers can then be assumed to have moved out of the country without reporting this. In order to find out the number of overcoverage, annual information based on administrative registers from other government agencies have been used to identify those without imprints. Registers that have been used contain information that origins from The Swedish Tax Agency (demographical and taxation data), Försäkringskassan (data from social insurance system), The Swedish National Agency for Education (data from public school system, publicly organized preschooling, school-age childcare and adult education), The Swedish Migration Agency (protection and asylum seekers) etc. etc.
The indicator shown above has been evaluated using decisions made by The Swedish Tax Agency to withdraw registration from the national registration as migrants. These decisions are preceded by a period of normally at least two years with attempts to reach the person. The register based indicator tends to miss people with small income, where income from not just the single person but the summarized income for the whole family is considered.

### 3.2 Suspected overcoverage

A new indicator, suspected overcoverage, has been constructed consisting of the population above plus persons whose family income is less than one price base amount.
Figure 2. Suspected overcoverage. Source: TPR, Statistics Sweden.

It is very likely that the group above consists of too many individuals and therefore we must find a way to decide if it is likely that a person belongs to the overcoverage or to the population.

### 3.3 Model estimated overcoverage

When deciding whether a person belongs to the overcoverage or to the population, a number of different indicators are used that can point out what group it is likely that a person belongs to. The indicators are created by a set of rules that are based on register data from a single year or from several following years. Some indicators use information from one or two years ahead and some use information from previous years. For example, the combination of a person that belongs to the suspected overcoverage and who also was graduated from university one or two of the previous years indicates that it is likely that the person has left the country after graduation without reporting it. Another indicator considers whether the person is subject to cross-border commuting. If so, it is most likely that the person belongs to the population. Because certain rules indicate overcoverage better than others, or the reverse, that the persons belong to the population, a weighted count is preferable.
For the last two years (2012-2013) the estimated value of the model is over-estimated because the indicators based on future observations are not fully available. For the first year (2000), certain indicators for previous years are missing, which causes an under-estimation of the overcoverage.

<table>
<thead>
<tr>
<th>Year</th>
<th>Model estimated overcoverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>(35 270)</td>
</tr>
<tr>
<td>2001</td>
<td>39 116</td>
</tr>
<tr>
<td>2002</td>
<td>39 322</td>
</tr>
<tr>
<td>2003</td>
<td>39 988</td>
</tr>
<tr>
<td>2004</td>
<td>39 688</td>
</tr>
<tr>
<td>2005</td>
<td>42 144</td>
</tr>
<tr>
<td>2006</td>
<td>40 704</td>
</tr>
<tr>
<td>2007</td>
<td>40 718</td>
</tr>
<tr>
<td>2008</td>
<td>45 525</td>
</tr>
<tr>
<td>2009</td>
<td>55 086</td>
</tr>
<tr>
<td>2010</td>
<td>59 786</td>
</tr>
<tr>
<td>2011</td>
<td>62 938</td>
</tr>
<tr>
<td>2012</td>
<td>(65 757)</td>
</tr>
<tr>
<td>2013</td>
<td>(73 371)</td>
</tr>
</tbody>
</table>

Figure 3. Model estimated overcoverage. Source: TPR, Statistics Sweden.

For the last two years (2012-2013) the estimated value of the model is over-estimated because the indicators based on future observations are not fully available. For the first year (2000), certain indicators for previous years are missing, which causes an under-estimation of the overcoverage.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of persons without imprints</th>
<th>Suspected overcoverage</th>
<th>Model estimated overcoverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>50 329</td>
<td>67 877</td>
<td>(35 270)</td>
</tr>
<tr>
<td>2001</td>
<td>52 842</td>
<td>70 800</td>
<td>39 116</td>
</tr>
<tr>
<td>2002</td>
<td>54 210</td>
<td>71 065</td>
<td>39 322</td>
</tr>
<tr>
<td>2003</td>
<td>56 134</td>
<td>74 947</td>
<td>39 988</td>
</tr>
<tr>
<td>2004</td>
<td>54 412</td>
<td>72 505</td>
<td>39 688</td>
</tr>
<tr>
<td>2005</td>
<td>56 840</td>
<td>75 535</td>
<td>42 144</td>
</tr>
<tr>
<td>2006</td>
<td>56 731</td>
<td>75 586</td>
<td>40 704</td>
</tr>
<tr>
<td>2007</td>
<td>60 046</td>
<td>79 332</td>
<td>40 718</td>
</tr>
<tr>
<td>2008</td>
<td>67 868</td>
<td>88 930</td>
<td>45 525</td>
</tr>
<tr>
<td>2009</td>
<td>78 644</td>
<td>102 843</td>
<td>55 086</td>
</tr>
<tr>
<td>2010</td>
<td>83 329</td>
<td>102 008</td>
<td>59 786</td>
</tr>
<tr>
<td>2011</td>
<td>85 205</td>
<td>110 529</td>
<td>62 938</td>
</tr>
<tr>
<td>2012</td>
<td>86 211</td>
<td>113 997</td>
<td>(65 757)</td>
</tr>
<tr>
<td>2013</td>
<td>88 311</td>
<td>116 684</td>
<td>(73 371)</td>
</tr>
</tbody>
</table>

Figure 4. Overcoverage. Source: TPR, Statistics Sweden.
The results from the model give an estimated value of overcoverage that is lower than the number of persons without imprints and about 55 percent of the suspected overcoverage.

There is a relationship between the raising number of overcoverage in the years of 2009 and 2010 and an increasing number of immigrants three to four years earlier.

The disadvantage with the proposed model is that the results is not available until 18 months after the reference time point, and also that the two last years are overestimated because some information is still not available. Because of this disadvantage, it was proposed that Statistics Sweden should create a regression model based on the model in order to estimate overcoverage for the current year of the Total Population Register. This has also been done in order to fulfill the obligations from EU to create an estimation of the total number of residents in Sweden.

### 3.4 Attributes of the overcoverage

<table>
<thead>
<tr>
<th>Year</th>
<th>Overcoverage</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>35 270</td>
<td>54,9</td>
<td>45,1</td>
</tr>
<tr>
<td>2001</td>
<td>39 116</td>
<td>55,4</td>
<td>44,6</td>
</tr>
<tr>
<td>2002</td>
<td>39 322</td>
<td>56,0</td>
<td>44,0</td>
</tr>
<tr>
<td>2003</td>
<td>39 988</td>
<td>56,1</td>
<td>43,9</td>
</tr>
<tr>
<td>2004</td>
<td>39 688</td>
<td>56,5</td>
<td>43,5</td>
</tr>
<tr>
<td>2005</td>
<td>42 144</td>
<td>56,9</td>
<td>43,1</td>
</tr>
<tr>
<td>2006</td>
<td>40 704</td>
<td>57,4</td>
<td>42,6</td>
</tr>
<tr>
<td>2007</td>
<td>40 718</td>
<td>58,2</td>
<td>41,8</td>
</tr>
<tr>
<td>2008</td>
<td>45 525</td>
<td>57,9</td>
<td>42,1</td>
</tr>
<tr>
<td>2009</td>
<td>55 086</td>
<td>58,3</td>
<td>41,7</td>
</tr>
<tr>
<td>2010</td>
<td>59 786</td>
<td>57,8</td>
<td>42,2</td>
</tr>
<tr>
<td>2011</td>
<td>62 938</td>
<td>58,1</td>
<td>41,9</td>
</tr>
<tr>
<td>2012</td>
<td>65 757</td>
<td>57,4</td>
<td>42,6</td>
</tr>
<tr>
<td>2013</td>
<td>73 371</td>
<td>58,2</td>
<td>41,8</td>
</tr>
</tbody>
</table>

Figure 5 Overcoverage by sex. Source: TPR, Statistics Sweden.

The overcoverage is dominated by men and the share of men has increased over the period that has been studied.
A majority of the overcoverage are born in countries outside the Nordic region and as the immigration increases the share also increase with two to three years of delay.

<table>
<thead>
<tr>
<th>Year</th>
<th>Overcoverage</th>
<th>Sweden</th>
<th>Nordic countries excluding Sweden</th>
<th>Europe excluding the Nordic countries</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>35 270</td>
<td>36,9</td>
<td>6,3</td>
<td>22,2</td>
<td>34,6</td>
</tr>
<tr>
<td>2001</td>
<td>39 116</td>
<td>37,9</td>
<td>6,6</td>
<td>21,6</td>
<td>33,9</td>
</tr>
<tr>
<td>2002</td>
<td>39 322</td>
<td>36,8</td>
<td>7,1</td>
<td>22,1</td>
<td>34,0</td>
</tr>
<tr>
<td>2003</td>
<td>39 988</td>
<td>35,3</td>
<td>7,4</td>
<td>23,2</td>
<td>34,1</td>
</tr>
<tr>
<td>2004</td>
<td>39 688</td>
<td>34,0</td>
<td>8,1</td>
<td>23,3</td>
<td>34,6</td>
</tr>
<tr>
<td>2005</td>
<td>42 144</td>
<td>33,1</td>
<td>8,0</td>
<td>23,2</td>
<td>35,6</td>
</tr>
<tr>
<td>2006</td>
<td>40 704</td>
<td>34,3</td>
<td>9,3</td>
<td>22,4</td>
<td>34,0</td>
</tr>
<tr>
<td>2007</td>
<td>40 718</td>
<td>35,3</td>
<td>10,2</td>
<td>22,1</td>
<td>32,4</td>
</tr>
<tr>
<td>2008</td>
<td>45 525</td>
<td>32,5</td>
<td>10,1</td>
<td>25,2</td>
<td>32,2</td>
</tr>
<tr>
<td>2009</td>
<td>55 086</td>
<td>30,0</td>
<td>8,3</td>
<td>28,0</td>
<td>33,8</td>
</tr>
<tr>
<td>2010</td>
<td>59 786</td>
<td>27,4</td>
<td>7,9</td>
<td>29,1</td>
<td>35,6</td>
</tr>
<tr>
<td>2011</td>
<td>62 938</td>
<td>26,7</td>
<td>6,9</td>
<td>29,4</td>
<td>37,0</td>
</tr>
<tr>
<td>2012</td>
<td>65 757</td>
<td>27,7</td>
<td>6,5</td>
<td>28,0</td>
<td>37,7</td>
</tr>
<tr>
<td>2013</td>
<td>73 371</td>
<td>27,8</td>
<td>5,3</td>
<td>27,0</td>
<td>39,9</td>
</tr>
</tbody>
</table>

Figure 6 Overcoverage by country of birth. Source: TPR, Statistics Sweden.

<table>
<thead>
<tr>
<th>Year</th>
<th>Overcoverage</th>
<th>0-24 years</th>
<th>25-34 years</th>
<th>35-49 years</th>
<th>50-64 years</th>
<th>64- years</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>35 270</td>
<td>23,8</td>
<td>27,1</td>
<td>31,2</td>
<td>11,5</td>
<td>6,4</td>
</tr>
<tr>
<td>2001</td>
<td>39 116</td>
<td>23,3</td>
<td>27,4</td>
<td>31,6</td>
<td>11,8</td>
<td>6,0</td>
</tr>
<tr>
<td>2002</td>
<td>39 322</td>
<td>22,7</td>
<td>27,9</td>
<td>32,1</td>
<td>11,9</td>
<td>5,4</td>
</tr>
<tr>
<td>2003</td>
<td>39 988</td>
<td>22,0</td>
<td>27,8</td>
<td>31,1</td>
<td>12,2</td>
<td>6,9</td>
</tr>
<tr>
<td>2004</td>
<td>39 688</td>
<td>21,1</td>
<td>28,4</td>
<td>30,5</td>
<td>12,8</td>
<td>7,2</td>
</tr>
<tr>
<td>2005</td>
<td>42 144</td>
<td>21,2</td>
<td>29,9</td>
<td>29,7</td>
<td>12,7</td>
<td>6,5</td>
</tr>
<tr>
<td>2006</td>
<td>40 704</td>
<td>21,5</td>
<td>29,8</td>
<td>29,2</td>
<td>13,0</td>
<td>6,6</td>
</tr>
<tr>
<td>2007</td>
<td>40 718</td>
<td>22,8</td>
<td>29,1</td>
<td>27,9</td>
<td>13,6</td>
<td>6,5</td>
</tr>
<tr>
<td>2008</td>
<td>45 525</td>
<td>22,4</td>
<td>31,3</td>
<td>27,2</td>
<td>13,0</td>
<td>6,1</td>
</tr>
<tr>
<td>2009</td>
<td>55 086</td>
<td>21,9</td>
<td>32,4</td>
<td>27,3</td>
<td>12,8</td>
<td>5,6</td>
</tr>
<tr>
<td>2010</td>
<td>59 786</td>
<td>21,9</td>
<td>34,3</td>
<td>25,4</td>
<td>12,4</td>
<td>6,0</td>
</tr>
<tr>
<td>2011</td>
<td>62 938</td>
<td>20,8</td>
<td>34,1</td>
<td>25,9</td>
<td>13,3</td>
<td>5,9</td>
</tr>
<tr>
<td>2012</td>
<td>65 757</td>
<td>20,4</td>
<td>33,0</td>
<td>26,1</td>
<td>14,7</td>
<td>5,9</td>
</tr>
<tr>
<td>2013</td>
<td>73 371</td>
<td>19,3</td>
<td>33,5</td>
<td>26,1</td>
<td>15,7</td>
<td>5,5</td>
</tr>
</tbody>
</table>

Figure 7 Overcoverage by age. Source: TPR, Statistics Sweden.
The age-distribution of the overcoverage is dominated by people in the younger middle-age groups which reflect the fact that it takes some years from immigration to emigration and thereafter at least one year until there is no imprints in the registers.

3.5 Why isn’t the register corrected?

Although we have the tool to predict who may be a subject of overcoverage, Statistics Sweden finds it inappropriate to exclude those persons from the TPR. It is the Swedish Tax Agency that is responsible for the population registration and if Statistics Sweden exclude them from the TPR, there will be a lack of consistency with statistics produced by other statistical producers that use the same source. Even though the prediction may be statistically relevant we also know that we miss some persons and that we sometimes point out individuals that belong to the population. The best way to avoid overcoverage is to help the Swedish Tax Agency to correct the register in its origin.

3.6 Controls of the population registration

The Swedish Tax Agency has been able to use information about the overcoverage to make the control of the population registration more efficient and the result has had an impact on the population statistics. An example is shown in the table below. Late 2014 the Tax Agency made a control concerning one of the groups that Statistics Sweden had pointed out as a group with high level of suspected overcoverage. This led to population decrease in some municipalities that normally, as in 2013, have population growth these months.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stockholm</td>
<td>2,274</td>
<td>687</td>
<td>574</td>
<td>2,404</td>
<td>-3</td>
<td>-388</td>
</tr>
<tr>
<td>Uppsala</td>
<td>376</td>
<td>127</td>
<td>57</td>
<td>343</td>
<td>-245</td>
<td>-53</td>
</tr>
</tbody>
</table>

Figure 8 Population changes. Source: TPR, Statistics Sweden.

3.7 The impact of overcoverage for statistics based on surveys or based on register

The impact of overcoverage in the population register differs between statistics based on surveys and statistics based on registers. In survey based statistics the overcoverage will be part of the non-respondents. The reason for this is simple, those who have left the country will not respond to the attempts made by Statistics Sweden. The consequence is that the overcoverage will be treated in the same way as the rest of the non-respondents. If the survey is about unemployment the
overcoverage will be assumed to have the same rate of unemployed as the rest of the non-respondents.

Figure 9 Schematic picture of a target population for a survey. Source: Statistics Sweden.

If we consider the same statistics based on registers it is normal to begin with those who are registered as unemployed by the Swedish Public Employment Service. The next step is to combine them with the TPR. Those who belong to the TPR and doesn’t have registration as being unemployed are considered as being not unemployed. Those who are not unemployed may be employed, students, retired etc. All of the overcoverage will be considered as being not unemployed.

Figure 10 Schematic picture of a register based population. Source: Statistics Sweden.
All statistics based on registers are normally done in the same way; a register with those who are subject to have a certain characteristic and the rest who are not subject to this characteristic. The latter group normally includes all of the overcoverage.

4. Conclusion

This work has increased the awareness in Statistics Sweden of the overcoverage and impact on statistics, both for statistics based on survey and register based statistics.

5. References

Övertäckning i Registret över totalbefolkningen – en registerstudie. Bakgrundsfakta 2015:1, SCB


Greijer, Å. (1997b). Skattning av övertäckningen i RTB med hjälp av AKU. Metodrapport från BoR-avdelningen 1997:12, SCB.


Qvist, J (1999). Täckningsproblem i RTB, R&D 1999:1, SCB


Skatteverket 2013. En analys om folkbokföringsfelet, dess kostnader och konsekvenser

Bengtsson, T., Berg, S. Så påverkas statistiken av övertäckning. Artikel Välfärd 1/2013, SCB