

Dealing with under coverage in administrative sources

Topic 1 – Bringing in information from where we can get it

Keywords: capture-recapture method, undercount, linkage correction

Introduction

If one uses administrative sources for official statistics, e.g. a population register or a combination of registers that could be used as a population register, one of the important quality issues is whether this source has under coverage. Under coverage is defined as register's number missing elements of the target population.

There are different ways to estimate the number of missing elements. The first is that the categories that are missed in the population register are defined, one looks for a source to count or estimate the size of each category and the intersection of the separate categories. The under coverage is then the total net number of missing elements that appear in these categories. However, the data sources are always imperfect, therefore only if you postulate a number of assumptions, one can come to an estimate (Bakker & Daas, 2012). The second method is that a post enumeration survey or a second register is linked to the population register and then apply capture-recapture models.

However, if this method is used you have five important assumptions:

- 1) the inclusion probabilities of both sources are independent;
- 2) all elements of the population should have a positive inclusion probability;
- 3) the population is closed;
- 4) there are no erroneous captures;
- 5) the sources are perfectly linked (Gerritse, et al. 2016; Di Consiglio & Tuoto, 2015)

Methods / Problem statement

What is the best way to estimate the under coverage of administrative sources, in particular population registers? We describe and apply the two methods mentioned in the introduction.

Results / Proposed solution

In the paper, we present both ways to estimate the undercount and show in what way the assumptions are met as well as possible. We apply both procedures to data of the Netherlands for 2010-2014 and present the results from both procedures. We will discuss the advantages and disadvantages of these procedures.

Conclusions

The results of both procedures lead to slightly different outcomes. Given the large amount of uncertainty, in particular caused by the assumptions that could not be verified in the data, applying both procedures leads to more knowledge of the under coverage. Both procedures are labour intensive and expensive. Therefore, it is not reasonable suitable to apply these in regular statistics production. However, it is necessary to investigate the under coverage of sources if one wants to use new administrative sources.