

Kåre Vassenden

## 9. Data basis for Statistics Norway's migration statistics

*Prerequisites for comprehensive and reliable statistics in the field of migration and immigrants («migration statistics») are the existence of relevant data, these data being available for the statistical office, and legal and technical opportunities for processing the data and making them feasible for the production of statistics.*

### 9.1. The data capture

In general, Statistics Norway obtains data for the production of person statistics from

- nationwide administrative registers
- institutions or agencies each covering limited geographical areas
- direct collection from the informants via interviews, postal surveys and complete counts.

The data collection from the registers and institutions is based on the right the Statistics Act gives Statistics Norway to exploit such data sources for the production of statistics.

The administrative register data form the basis for drawing samples, and are subsequently linked to supplement the survey data. In this way the access to good register data is a basic prerequisite for all Norwegian migration statistics, even those that are based on surveys.

In addition to the data retrieved externally, Statistics Norway has considerable

amounts of data from earlier data collections. These data also represent an important source for the production of statistics.

Statistics referring to immigrants are mainly based on counting people. In other cases the unit is events or circumstances related to persons. Almost all accessible data for such statistics are at individual level.

### 9.2. The most important ID number series

Most of the individual data that Statistics Norway receives has a Personal Identification Number (PIN) as identification. However, the so-called D-number is also used in certain cases, and other number series are used in relation to data from the Immigrant Administration.

PINs are assigned by the Population Registry and registered in the Central Population Register (CPR) database. The population registration system is a part of the Tax Administration. Everyone born in Norway receives a PIN, as do immigrants that meet the conditions for being registered as resident in Norway. In addition, some

smaller categories are given a PIN even if they have never resided in Norway (e.g. children of Norwegian citizens abroad).

At the beginning of 2008 there were almost 7.4 million PINs in the CPR, covering 7.16 million people. In addition to the 4.7 million residents at that time, the CPR contained almost 2 million deceased persons and 440 000 emigrants that have ever resided in Norway since the establishment of the register in 1964. In addition to PINs, the CPR contains of course information related to the population registration.

The D-number is assigned to persons that do not qualify for a PIN, but nevertheless have economic relations with the Norwegian authorities. The Tax Administration is also responsible for this number, and the D-number register forms part of the CPR database. More than 1 million D-numbers have been assigned through the years. The D-number has the same format as the PIN, with the exception of the date of birth (the two first digits), which is increased by 40. If a person that only has a D-number immigrates to Norway (i.e. is registered in the population register as resident), a link from the D-number to the PIN is established. In many ways, the PIN and the D-number belong to one and the same ID number series. Otherwise, very little information is connected to the D-numbers.

In the Norwegian Directorate of Immigration, a DUF number is assigned to almost everyone that is registered in the Aliens Register (UDB). For those with both a DUF number and a PIN (or D-number), there are arrangements in place to ensure that this connection is stored in both the UDB and CPR. The coverage of old registrations, however, is not as good as could be hoped.

### 9.3. Processing the data

The data basis for immigrant statistics is normally only one part of the data basis used for the production of the statistics for the total population, but some data sources are particular to persons with an immigrant background.

When data are received in Statistics Norway they are subject to extensive processing before they can be utilised in the production of statistics and/or as input for other data development projects. Data are checked and repaired or improved in different ways during the processing stage. New variables that are more suitable for the statistics are often produced. Information from previous editions or other sources are often matched up.

The quality can be improved when utilising data from several sources. Matching is also necessary when data from different fields shall be merged. When combining data in new ways, new knowledge can emerge. Extensive matching between different sources can often lead to difficulties in selecting only one specific source for a certain processed variable or complete data set.

The PIN has been a key in the production and development of social statistics in Norway. The existence of this ID number series in practically all relevant input registers enables the simple and secure forming of links.

In Statistics Norway, the base variables are produced in one place and then distributed to the subject areas that need them in their production of statistics. There the variables are linked by means of the PIN. For example, who are classed as immigrants is decided one place, the level of education another place, and income a third.

This exchange of data between subject areas and the preparation of data for research projects mean that in principle there are two objectives with data production in Statistics Norway. The data shall be suitable for both the direct production of statistics and as input to other data development processes. All data in Statistics Norway eventually becomes different kinds of statistics, but prior to that happening, data are needed as individual level data.

From a legal perspective, data from administrative primary registers and separate collections become «statistical registers» as soon as they are received in Statistics Norway. This means that these data shall only be used for the production of statistical data and statistics. The use of registers in Statistics Norway is governed by the Statistics Act and the Personal Data Act.

Technically, registers and variables do not become statistical until they have been through a preparation process that turns them into something other than the original, administrative data.

#### **9.4. The Norwegian register society is integrated and comprehensive**

What characterises Norway and some other countries (primarily the Nordic countries) with regard to administrative register sources, is that they cover a relatively large part of society at large, that the quality of the registers is relatively high, and that there is contact between registers that enables a certain degree of coordination. Additionally in the Norwegian society, and not only internally in Statistics Norway, data are exchanged between registers that are responsible for each administration area.

The key role of the CPR for personal data in Statistics Norway has already been

mentioned. For migration statistics, data from the Aliens Register and the Adoption Register are significant, both directly and indirectly (via the CPR).

The education statistics are partly based on administrative systems, partly on data directly from the educational establishments and otherwise from bodies such as the Norwegian State Educational Loan Fund and the Norwegian Welfare and Labour Administration. The latter agency, with its Employees Register, is the most important supplier for the labour market statistics, which also utilises the Tax Return Register and the Register of End-of-the-Year Certificates. The most important sources for the income statistics are various registers in the Tax Administration.

Several other administrative registers are also used in Statistics Norway.

#### **9.5. The sources for the classification of persons by immigrant background**

In Statistics Norway, what is classified as immigration and who have immigrant backgrounds is defined in the population statistics system. Central immigrant statistics variables are citizenship, change of citizenship, country of birth, immigrant category and country background, reason for immigration and first date of immigration.

As the source for these variables, notifications received from the CPR are mainly used, but data on country of origin from the Population and Housing Census 1970, for example, are used to supplement the information from the population registration system.

Data from the Aliens Register is used to produce the reason for immigration variable, in addition to demographic data in Statistics Norway.

From the «core» of migration statistics, represented by files with country of birth and other population data, individual level data are forwarded to subject areas such as education, population census, health, national insurance, crime, labour market, income, interview and so on, for linking to the data files of these subject areas.