

# **Deriving Educational Attainment by combining data from administrative sources and sample surveys**

Recent developments towards the Census of 2011

**Frank Linder and Dominique van Roon, Statistics Netherlands**

Key words: Accuracy; administrative register; bootstrap-method; consistent estimation; educational attainment; micro-integration; micro-linkage; Netherlands; population census; sample survey; sampling design; Social Statistical Database; weighting strategy.

Data on educational attainment are crucially important in socio-economic research and government policy. Multivariate analyses often show that education has an undeniable impact in explaining a variety of social phenomena. For example, prospects on the labour-market are generally better for those who are well educated. That knowledge may convince the government of the need to prevent dropping out of school and to invest more in education programmes. Also, several studies indicate that higher levels of education tend to generate more average income. Altogether, it is not surprising that educational attainment is a standard variable in the Census Programme.

The Virtual Census of 2001 in the Netherlands used the Social Statistical Database (SSD) as its key source. The SSD can be considered as a statistical framework that provides information on demographic and socio-economic issues. It is constructed by micro-linkage and micro-integration of several administrative registers and household sample surveys, which ensures coherence, consistency and completeness of the SSD data. Whenever administrative data sources are available, the SSD gives priority to these. Sample surveys are explored to compensate for information that is not (yet) in registers. At the time of the 2001 Census the only source providing data on educational attainment was the sample of the Labour Force Survey (LFS).

In the last decade a wide variety of administrative education registers has come at the disposal of Statistics Netherlands. It has increased the information content in the field of education considerably. So, from now on it is possible to produce more reliable estimations on education level than the LFS-results from earlier years, in particular when smaller subpopulations are involved. It is a development of which the next Census of 2011, with its detailed table programme, will definitely take advantage of.

As the education registers do not cover the entire Dutch population, the LFS still plays an important part to fill the gaps. The education career of most older citizens took place before registers came into consideration for official statistics. Besides, information on studies abroad is not found in education registers. In these cases the LFS compensates for the lack of information in the administrative sources.

Deriving educational attainment by combining data from registers and the LFS supplement is far from simple. Problems, such as the complex sampling design for the combined register and sample data, have been solved by applying sophisticated methodology. At present the weighting strategy is under consideration for revision. In addition bootstrap-methods are being developed for the construction of variance estimators and confidence intervals, to determine whether the estimation of education levels is accurate enough for dissemination. The results of this research will be incorporated in the production process before the actual data preparation for the Dutch Virtual Census 2011 is due to start.