

Elisabetta Vassenden

The register of the level of education:

a product of the integration of other registers, census and surveys

Work Session on registers and administrative records
for social and demographic statistics
Geneva, 23-25 January 1995

**THE REGISTER OF THE LEVEL OF EDUCATION:
A PRODUCT OF THE INTEGRATION
OF OTHER REGISTERS, CENSUS AND SURVEYS**

by Elisabetta Vassenden

23 November 1994

The register of the level of education: a product of the integration of other registers, census and surveys

1. Good conditions to develop individual statistics

Norway is a country where two important factors facilitate the development of individual statistics. First and foremost, the Central Population Register with the National Identification Number; secondly, the Statistics Act of 1989, that makes it easier for Statistics Norway to utilize data from public systems established for administrative purposes.

The Central Population Register (CPR) was established in 1964. Its aim was to favour both statistical and administrative purposes. The main pillar of the CPR is the National Identification Number, a unique and permanent 11-digit number that identifies each person that ever lived in Norway since 1 October 1964. This individual number system is now extensively used by both governmental authorities and private organizations. It is also used as a unique identification in many registers, and as a linking key when registers are matched.

According to the Statistics Act of 1989, Statistics Norway has an official role in the planning and establishment of administrative data systems. Authorities and private organizations must keep Statistics Norway informed about their plans. Statistics Norway in its turn has the right to influence systems that may be of use for statistical purposes.

The register of the level of education of the population was created out of these conditions, integrating several sources with the specific purpose to meet the requirements of the users.

2. Importance of the level of education

The level of education has always been an important factor in the background of a population. It is one of the strongest components in social background, and the 'amount of knowledge' in the country can usefully be compared to what use the knowledge is actually put to. Since the level of education shows considerable variations between different regions (figure 1), both governmental and local authorities are interested in these data, and they use them actively in administrative planning. (See figure 2 for education level in some European countries.)

Statistics about the level of education were produced in the 1950, 1960, 1970 and 1980 census. This was not often enough to meet the needs of external users and of other register holders within Statistics Norway, so a regular register was established and updated every year from 1985. Today this register is not only used extensively on its own to produce education data, but it is matched with almost all individual-based statistics from Statistics Norway, both regular and surveys. Most surveys now don't need to ask the informant about her own education any more, because it is time- and money-saving to extract this information from other official statistics.

3. Contents of the register

The register of the level of education is one of the final products of individual-based education statistics. It is updated once a year. The register comprises the population resident at 1 October, aged 16 and older at 31 December. The 1993 register contained 3.4 million records, and was organized as a sequential file in a mainframe environment.

In addition to the National Identification Number, necessary as a linking key in building and updating the statistics, the register contains an essential core of demographic and education data: age, sex and municipality of residence fall into the first group, while highest completed education, completion date, country of completion and education in progress fall into the second group. Additional information was added to the essential demographic core when the users requested data about immigrants: nationality, country of birth and date of 1st immigration.

In a future implementation we will also add a new and more accurate immigrant classification to identify different immigrant types (e.g. second-generation immigrants or adopted persons) and reason of entry. The parents' level of education and the subject's social background will also be linked to the original file.

Main variables. Overview

Linking key	National Identification Number
Demographic core	Age Sex Municipality of residence
Education core	Highest completed education Completion date Country of completion Education in progress
Immigrant data	Nationality Country of birth Date of 1st immigration
Future implementation	Immigrant type Reason of entry Mother's highest completed education Father's highest completed education Social background

4. Actual population

The CPR is the original source of the actual population's data, extensively checked and processed by Statistics Norway in the Population Statistics System. This is an effective production system for population statistics that secures data for use in other individual statistics. The core of this system is a situation file (stock data) and several files on transactions (notification registers) based on regular input from the CPR. The situation at 1 January can be adjourned at any date during the year using the transaction files.

5. Construction of the register

In the life of the register of the level of education it is possible to identify three main phases. In the first one the register was a census implementation supplying important information obtained at the lowest possible cost. The second phase opened the way to regular statistics, while the third one concentrates on improving the data quality and filling gaps.

5.1. First phase: implementing the census

Individual data on the level of education was collected in the 1970 census, and in the early 1970s a complete reorganization of education statistics was carried out, making individual data the basis for the regular statistics. This paved the way for a more flexible system, providing more detailed information on the education system. One of the main advantages was the possibility of updating the census data with information on education completed in Norway supplied by the regular individual statistics. This was done to provide the 1980 census with a practically ready-made register of the level of education.

The only data the register lacked was information on education completed abroad after 1970. The solution to this problem was an additional question in the 1980 census, asking for this type of education only. Also this information was used to update the register.

5.2. Second phase: regular statistics

The second phase started in the late eighties. At that time the register was established as regular statistics produced every year. The first regular register is from 1985. The original plan was to supply statistics for *all* the years following the census, but there were technical problems in finding the actual population at 1 October for 1981-1984. These problems were due to the reorganization of the CPR from batch-oriented files to a database. All the data from the old CPR was loaded into the new base in 1985, making it impossible to reconstruct the actual population living in Norway at such an unusual date as 1 October *before* 1985.

To construct and update the register (figure 3), all known data on completed education is assembled together into a comprehensive file, not taking into consideration whether the person in question is alive or dead, living in Norway or abroad. When the sources give more than one completed education for each person, a special procedure compares the education activities in question and chooses the highest one. The comprehensive file was established by collecting together data from the 1970 and 1980 census, plus information on education completed in Norway during the school years 80/81 to 84/85.

The comprehensive file is not built up from scratch every year, but only updated with education completed in Norway during the last school year (figure 4). When the file has been updated with this year's completed education, it is matched to the actual population living in Norway at 1 October to obtain this year's updated register on the level of education.

5.3. Third phase: filling gaps

The third phase in the life of the register is still in progress. It consists of filling gaps and improving the data quality.

The register's most serious shortcoming was (and still is) the lack of information on education completed abroad after 1980. The problem is extremely evident with people born abroad: during the late eighties and the nineties a lot of refugees and asylum seekers came to Norway. Unless they complete some kind of education in the country, they have 'Unknown education level' in the register because there are no administrative sources from which their foreign education can be extracted. In 1989 about 54 per cent of persons born abroad had 'Unknown education level'. This worried us and the register users a lot, since 'foreigners' (broadly speaking) are a very interesting group for both governmental authorities, public administrative bodies and mass media.

5.3.1. The survey 'Education completed abroad'

Something drastic had to be done to preserve the value of the register. We sent a questionnaire to all foreign-born persons that had moved to Norway for the first time between 1980 and 1990, aged 16 and older at the time of migration. The name of this survey was 'Education completed abroad'.

The informant had to classify her own education completed abroad according to the main criteria of the Norwegian Standard Classification of Education. This saved us a lot of manual coding, and the questionnaire was read optically. Sixty thousand questionnaires were sent, and net response rate was 70 per cent, while 3 per cent of the total was returned because the actual person wasn't residing in Norway any more. This fact led to the awareness that routines for the notification of emigrations are not always followed, and that the CPR must

contain too many foreign-born persons as living in Norway. Migration statistics was involved in the problem, and examined the records of these persons. The CPR was notified, the Local Population Registries made thorough checks, and a side-product of 'Education completed abroad' were extra notifications of emigration for about 3300 people.

The results of the survey were used to update the register of the level of education. In 1989 the percentage of people born abroad with unknown education was reduced from 54 to 27.

However, the problem is not solved forever. We must have another survey shortly before the 2000 census, but we are also hoping to find a way to collect data regularly, preferably in connection with immigration notifications. The question is still open.

5.3.2. A hidden shortcoming: education completed abroad for people born in Norway

Very few persons born in Norway have 'Unknown education level'. However, we are aware that more and more Norwegians go abroad to study each year, mostly at tertiary level. This means that the level of education is underestimated for a part of the population.

To solve this problem, we now collect data from the State Educational Loan Fund. The Fund gives loans and/or scholarships for studies abroad to eligible students, the majority of whom are Norwegian born. The Fund has no exact knowledge of whether the students complete their education or not, but they believe that when the students give notice that they have 'finished' with their education, over 95 per cent of them have completed it, too.

A student who has 'finished' her own studies must pay back her loan, so to update the level of education Statistics Norway uses the Fund's 'Repayer register', where we can find information about the country and type of education (probably) completed.

We started using this administrative register only recently, but we are aware of quality problems that must be further looked into. Moreover, the Fund's coding of education is sometimes sketchy, following only one criterion of the Norwegian Standard Classification of Education. It is then imperative to co-operate with them, to help them use the Standard and generally raise the data quality.

5.3.3. Foreign schools in Norway

Examining data on education in progress, an alarmingly low enrolment rate was found for pupils born in western countries. This led to the discovery of a few foreign schools we didn't know about, mostly in Oslo and Bergen, with a majority of pupils from western countries.

These schools are private, but they do not receive financial support from the state, as most of the private schools do. Because of this, neither the Ministry of Education nor the municipal education authorities were properly aware of the existence of these schools.... We will of course include them in our statistics from next year.

6. Many sources in education statistics

To produce statistics on completed education it is necessary to use several sources. Also information about enrolment has consequences for completed education, because when we know that a pupil is enrolled, later we can ask the school if the pupil has completed the education in question.

Statistics Norway collects data both from several administrative data systems and from many schools. The main administrative sources are the County pupil enrolment registers, the Journeyman contract register and a register from the Directorate of Labour.

Each county has a County pupil enrolment register to regulate the enrolment of pupils in education at second level, second stage. In the majority of cases the pupils don't apply directly to the schools, but centrally to the county's enrolment system. Although the organization of the registers may vary from county to county, the output system is the same, to ensure that we receive the same kind of data from each unit. Additional variables have been added to the registers to identify the individuals we need in our statistics.

The Journeyman contract register contains information about contracts in progress and completed contracts. It was established and it is maintained by a private IT-company on behalf of the Ministry of Education.

From the Directorate of Labour we collect information on courses for unemployed, comprising both pupils and completed education.

Centralized administrative systems, spanning over as many schools as possible, are preferred as sources. Unfortunately, these systems are not enough to supply information on all schools, so that quite a lot of data must be collected directly from each school. A widespread data capture, with over 1000 different sources, is perhaps the main difficulty in the production of education statistics.

Most of the data is transferred electronically to Statistics Norway by tape, cassette, diskette or telecommunications. Some data is still collected with forms and punched. We are striving to abolish all manual work, using new administrative sources and persuading reluctant schools to transfer data electronically. This is neither an easy nor a quick job. After collection, all data is processed with extensive checks and corrections, and missing information is provided.

7. Prerequisites to produce statistics based on individual registers

I have previously pointed out the importance of a national identification to construct a statistical system based on individual administrative registers, but this is not the only prerequisite needed to develop such a system.

It is also necessary to have (many!) good classification standards with solid and logical structures. Each standard must be used to classify only ONE characteristic. As an example I can mention ISCED, the International Standard Classification of Education, used to classify educational programmes. To classify delivery systems/school types it is necessary to use another standard. The worst thing one could do, would be to 'mix' school characteristics with programme characteristics. (ISCED is actually being revised now, and this hair-raising proposal is being discussed.) A good standard has a good structure, and good standards do not overlap. These qualities will allow for more flexibility in an individual data system.

A statistician dealing with individual registers must be prepared to play many roles to optimize the data capture. She has to be a detective to trace unknown and obscure sources, exploring further possibilities and opening for rationalization within the system. She must be tireless in following up the developments of the administrative sources. In 99 per cent of the cases there are bound to be problems with the sources all the time. 'Suspicious', 'ceaselessly seeking hidden faults' and 'always pursuing' are qualities well-suited to our kind of statistician.

Last but not least, in every individual-based system personal security must be regulated by strict laws. Professional secrecy is essential to secure data protection. The structure around the data systems must guarantee the security and protection of the individual. Good individual statistics cannot flourish where the public mistrust the system.

Figure 1.

Percentage of the population aged 16 and older with tertiary education. Municipalities with highest and lowest percentage. 1 October 1993

Per cent

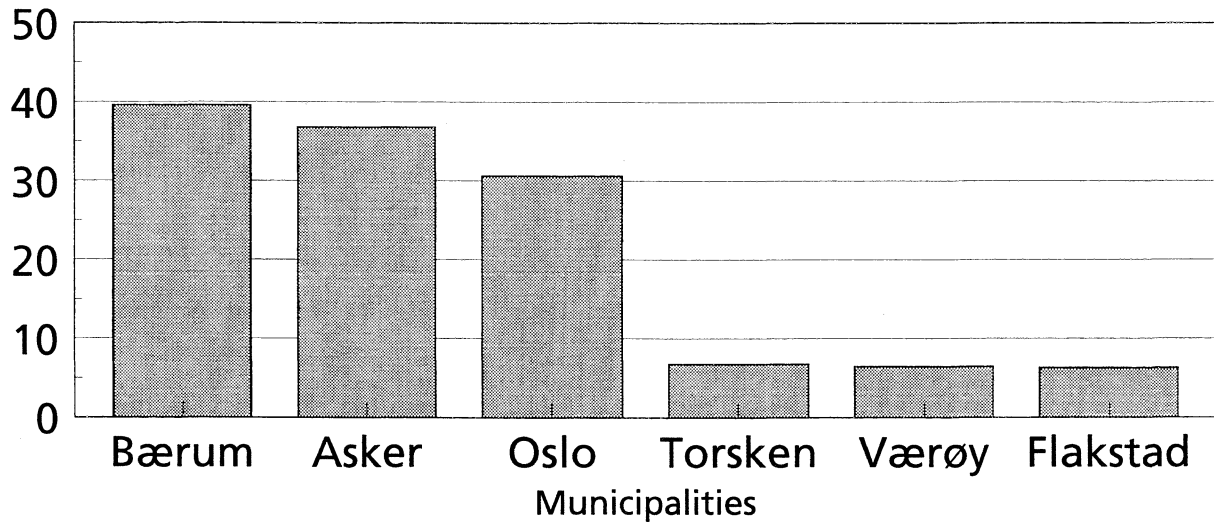


Figure 2.

Level of education for persons aged 25-64 in some European countries. 1991

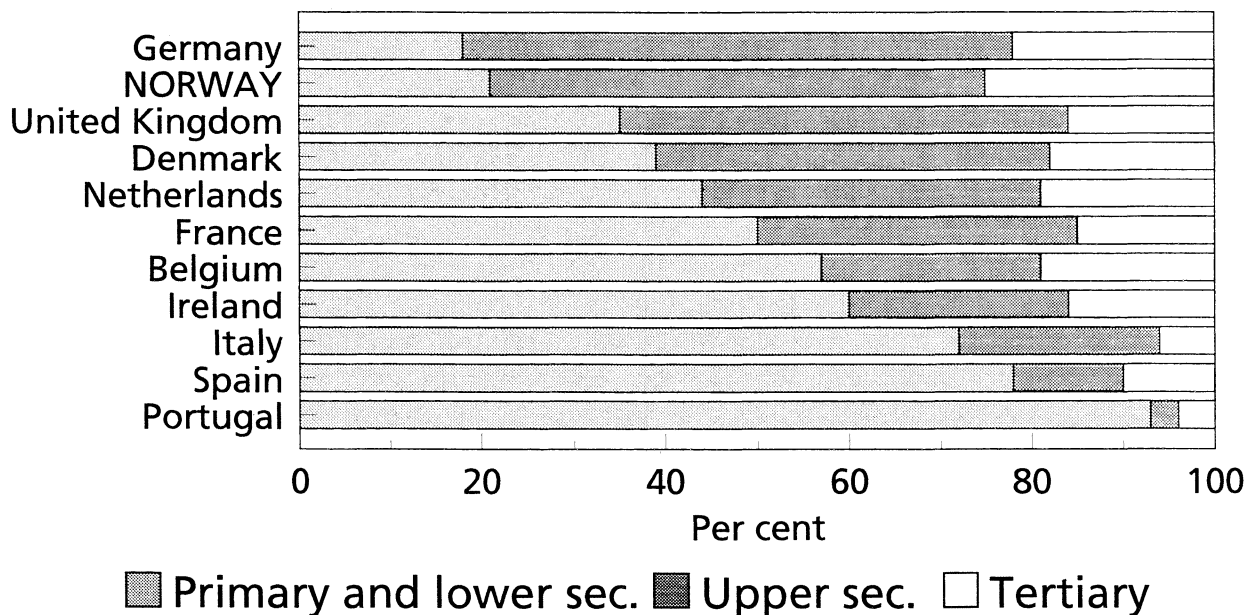


Figure 3.

DATA FLOW

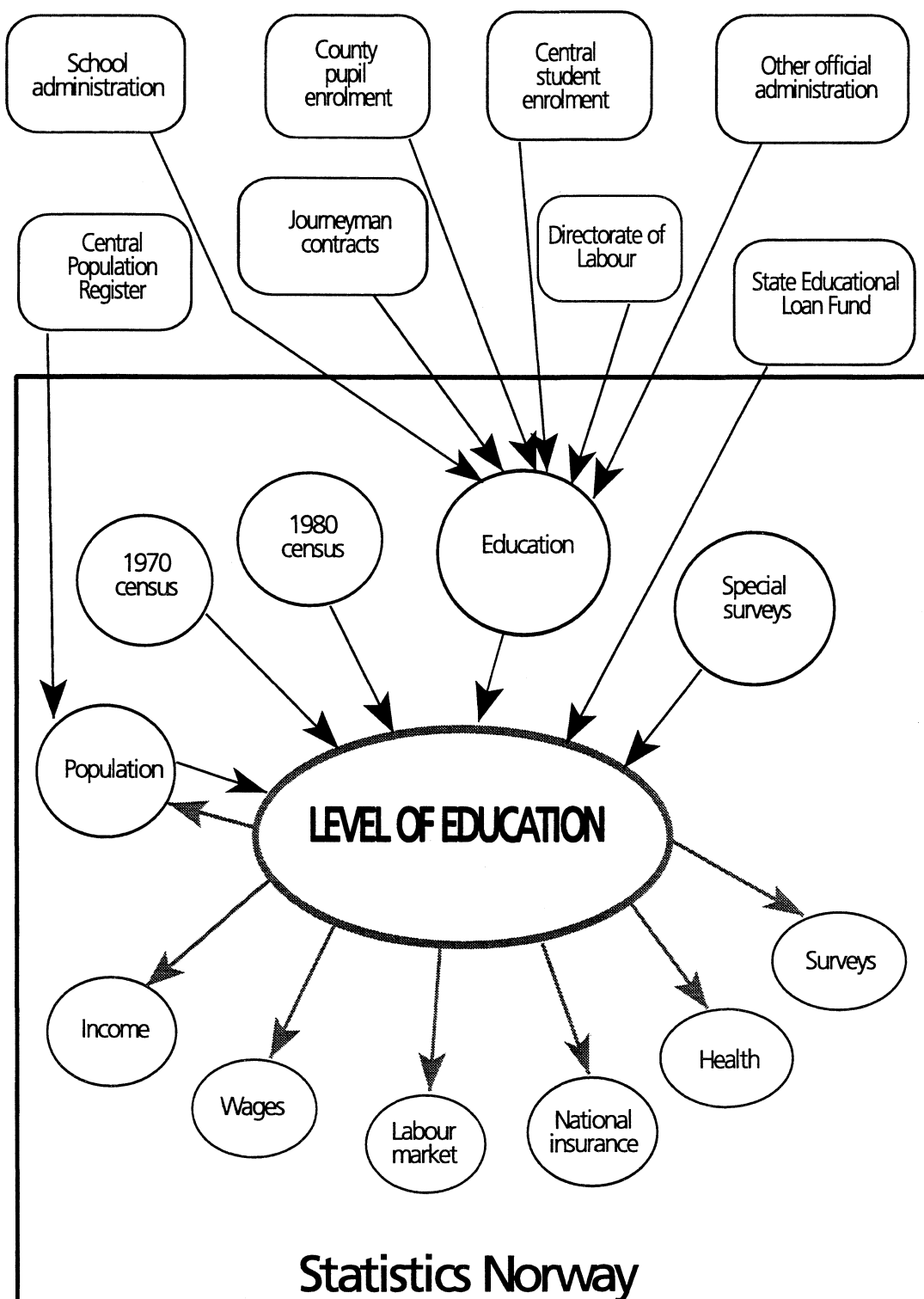
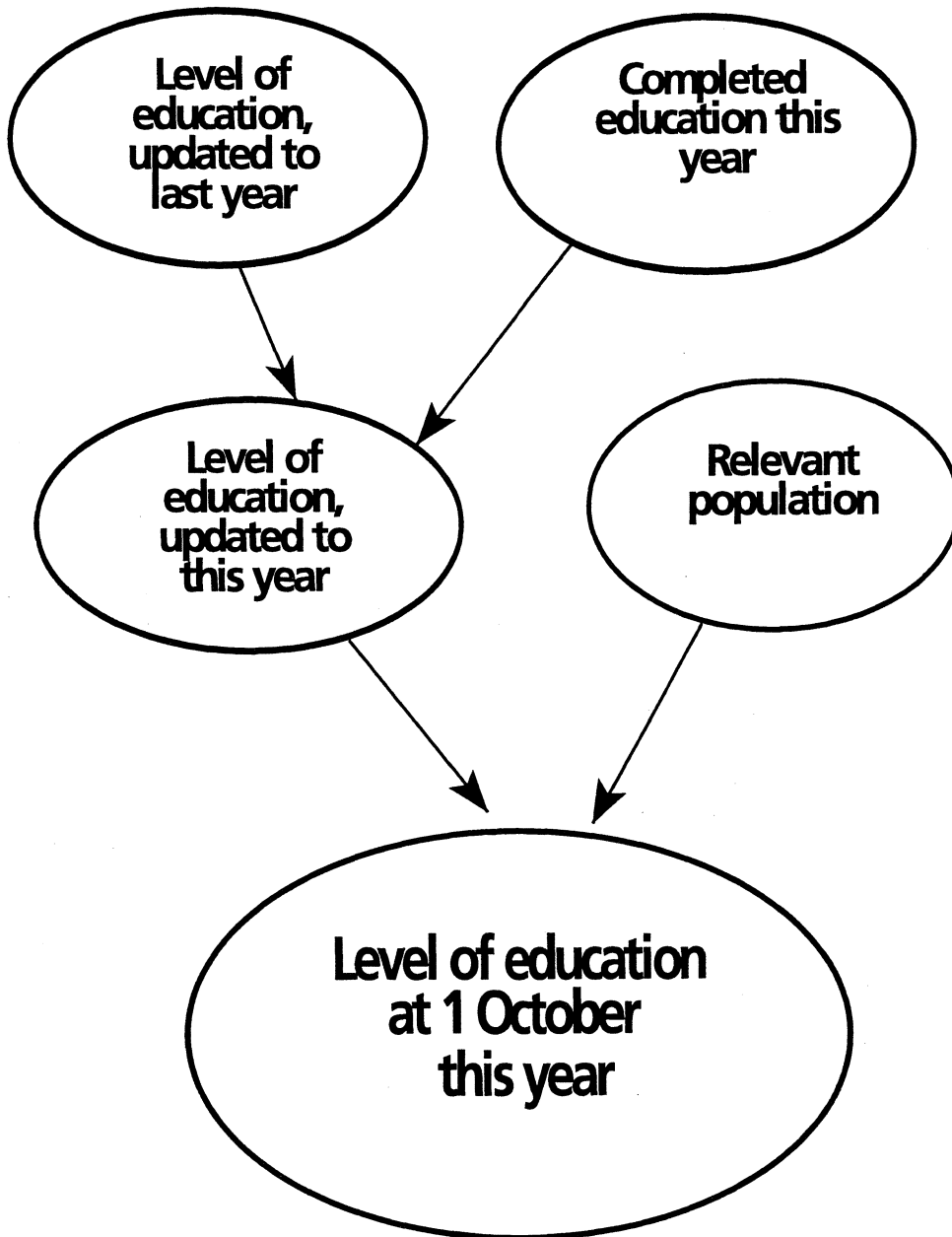


Figure 4.

SIMPLIFIED REGULAR UPDATE



Statistisk sentralbyrå

Oslo
Postboks 8131 Dep.
0033 Oslo

Tlf.: 22 86 45 00
Fax: 22 86 49 73

Kongsvinger
Postboks 1260
2201 Kongsvinger

Tlf.: 62 88 50 00
Fax. 62 88 50 30



Statistisk sentralbyrå
Statistics Norway