A roadmap to EU comparable data on mortality differentials by educational attainment

Veronica Corsini, European Commission - Eurostat

All EU Member States are facing substantial inequalities in the health outcomes between different sections of their population based on socio-economic status (SES). While overall levels of mortality have declined in all groups, experts note that relative mortality differences between high and low SES groups have remained unchanged or even increased. Therefore, Eurostat, as the Statistical Office of the European Communities, has been requested to develop comparable information on mortality by socio-economic status on a regular basis for all EU Member States.

One way of obtaining such data would be to ask the indicators directly to the EU Member States. However, data on mortality and on SES group in the countries differ considerably in terms of completeness, coverage, accessibility, quality and possibilities to link records of citizens in different databases. Moreover, definitions, approaches and measures may differ when mortality statistics by SES are produced, if any. Some countries informed Eurostat about past studies on the topic and the related data quality and data comparability issues. For example, Italy, Lithuania, Austria, Norway and Switzerland have used a linkage between national censuses and annually collected death record data and with education as socio-economic characteristic. Such studies were ad-hoc exercises carried out on the occasion of past census rounds. A limited number of countries, for example Denmark, France, the United Kingdom, calculate more or less regularly life expectancy by SES.

Given the heterogeneity and irregular availability of such indicators in the Member States, Eurostat decided to collect the needed input data in order to subsequently calculate mortality differentials by SES using the same methodology for all available countries, obtaining then comparable indicators. Several decisions were then necessary. First, different indicators of socio-economic status can be proposed, each showing advantages and limitations. The three main indicators are educational attainment, occupational status and economic status (income and wealth status): <u>educational attainment</u> has been chosen by Eurostat because, provided a standard classification is used, it can be determined for all individuals (therefore, lower non response is expected), it is more stable and it is in line with countries' practices.

Second, concerning mortality, several indicators can be proposed which are usually death rates or life expectancy and life table based information: therefore, information on mortality and on population at risk with the same breakdown by educational attainment is needed in order to compute the requested indicators. The choice Eurostat has made goes towards <u>life expectancy</u> because it has the advantage of ready interpretation and in addition, it is comparable across populations with different age structure.

Third, it is known that two main approaches for obtaining the data can be distinguished:

   i. *cross-sectional, unlinked studies*: information on deaths, including SES of deceased, is coming from death certificates while population data, including their SES, is coming from the census or registers or census based population estimates (cohort-component). The most serious limitation of this approach is the lack of comparability of the SES information on the death certificate (reported by a proxy informant, usually a relative of the deceased person) with that of the census (self-reported by the person): this is the so-called numerator-denominator bias. It has been found that census-unlinked mortality estimates lead to bias of the mortality indicators for some socio-

economic groups. Despite this bias this approach has the advantage of being relatively cheap to implement and more readily available.

ii. *prospective, linked record studies*: the chosen socio-economic characteristics of the deceased persons are obtained by linking the death certificates with census information or population registers. The linkage is done through a unique, personal identity number. Therefore, the same source is used to derive SES information for the numerator (death registers) and denominator (census/population register) of the rates by SES. Death information is followed for a certain period after the date of the census, for example half a year or several years. This is the approach to be preferred because individual data linkage avoids inconsistencies between the numerator and the denominator. By linking, a bias is to be expected because the SES is measured at the time of the census and not at the time of death. Anyway this bias may be considered much smaller than the numerator/denominator bias if the length of time between the census and the occurrence of death is short. If the linking cannot be done via a unique personal number (personal identifier not available or national legislation preventing its use), deterministic methods (using a combination of personal identifiers like sex, date of birth, place of birth, marital status, place of residence, etc.) or probabilistic methods (based on decision rules on the likelihood that two records in death and population belong to the same person) should be used instead. Variants can be proposed, based for example on a sample of the population taken on the occasion of the census and then linked to death records.

As a first step, it has been decided that Eurostat should calculate the mortality differentials using data already available and collected in the yearly demographic questionnaire. Since the reference year 2007, Eurostat requests countries to annually transmit a breakdown of population and number of deaths by ISCED97 broad education classes (ISCED0, 1 and 2; ISCED3 and 4; ISCED5 and 6; unknown; not applicable). The calculations of such mortality differentials can be done directly for the countries providing both deaths and population series by the requested breakdown. For the countries not providing breakdown for the population at risk, alternative data sources are being evaluated: one source could be the European Union Labour Force Survey (EU LFS), from which relevant information on educational attainment could be indirectly used in order to estimate the needed breakdown for the population.

This cross-sectional approach has the advantage of a quick feasibility; among its drawbacks, it is clear that it can be done only for those countries sending at least series on deaths by sex, age and educational attainment. A numerator/denominator bias should also be expected for countries with non-register based population estimates and for those countries where alternative populations would be used. Methodological discussions are currently on-going on these biases, in particular on the treatment of the categories "unknown" and "not applicable" of the ISCED97 classification: a sensitivity analysis has been recently carried out to identify which approach to follow among the three possible

- "unknown" and "not applicable" are ignored;
- "unknown" and "not applicable" are included into category ISCED0_2;
- "unknown" and "not applicable" are redistributed into categories ISCED0_2, ISCED3_4, ISCED5_6 proportionally to their relative sizes by sex and age.

Another point for discussion concerns the age censoring: due to data unavailability for ISCED groups at higher ages and to the limitation of the LFS data, the relevant age range by ISCED could be limited to 20 to 74 years old; mortality rates up to 20 and from 75 to 85+ years old could be considered equal to those of the total population; the consequence of using mortality rates referring to total population for age bands above 74 would be underestimation of differences by socio-economic group.

The cross-sectional approach is currently feasible for 12 EU Member States but Eurostat is aware that this number will increase in the near future due to recent developments in data availability.

In parallel, Eurostat is encouraging the National Statistical Institutes to carry out the "census records linkage" and follow-up on the occasion of the 2011 census round. Therefore, a list of guidelines and of related methodological issues is being developed on order to help the NSIs interested in doing this exercise. These guidelines refer to
- linkage procedure and software;
- handling of non-matched deaths;
- calculation of deaths rates;
- follow-up period;
- possible sources of bias.
Although it is likely that the requested data would not be available before 2-3 years after the census, immediate action is required. Despite the long implementation time the aim is to achieve a complete coverage of the countries taking advantage of the forthcoming census.
Based on the information Eurostat recently received from the countries, some NSIs are indeed planning to carry out the census linkage exercise to derive data on mortality by educational attainment; in some other cases, it would not be a classical linkage census - deaths but different registers would be linked to obtain the needed data.

Finally, discussion is on-going concerning migration, which can be considered a perturbing factor when studying trends in life expectancy by educational attainment. Eurostat intends to produce EU comparable results concerning mortality differentials for the total population at national level. Therefore, from the methodological point of view it needs to be considered how to follow immigration and emigration and the impact in the population at risk.