Estimating Differential Mortality from EU-SILC Longitudinal Data

A Feasibility Study

FACTAGE – WP 4
Deliverable 4.1

Johannes Klotz and Tobias Göllner

Statistics Austria

October 2017

Preface

This report was written by Senior Researchers of Statistics Austria, which is one of five members of the FACTAGE project consortium. FACTAGE is a project funded in the framework of the first call of the Joint Programming Initiative 'More Years, Better Lives' ('Extended Working Life and its Interaction with Health, Wellbeing and beyond'). FACTAGE is funded nationally by the Austrian Federal Ministry of Science, Research and Economics.

This report is the first in a series of three reports to be produced in FACTAGE Work Package 4. Its subject is the feasibility of estimating differential mortality figures from EU-SILC longitudinal data in a comparative European perspective. A second report will follow in early 2018 containing results, and finally a detailed technical report is to be produced in mid-2018.

This report was reviewed internally (within the FACTAGE project consortium) by the University of the Basque Country and the Centre for European Policy Studies as well as externally by Prof. Dr. Johan Mackenbach (Erasmus MC Rotterdam) and Prof. Dr. Franz Kolland (University of Vienna). Of course, any errors are solely the authors' responsibility.

Abstract

Socio-economic differences in mortality have become increasingly important in an era of pension reforms. Some European countries cannot provide any figures on the subject, and available figures are not easily comparable between countries because of different data sources, time periods and stratification variables. We present a new and relatively easy approach to obtain comparative European figures based on harmonized survey sample data. Longitudinal information of the EU-SILC survey (micro data on individuals and households) is extracted from Eurostat's User Database (UDB) which is available to researchers carrying out statistical analyses for scientific purposes.

Firstly, we discuss general pros and cons of longitudinal sample survey data for differential mortality analyses as well as specific EU-SILC UDB problems in coverage and content. We then describe in detail the necessary UDB data query and the following data editing for mortality analyses. We discuss some data quality indicators and present some exemplary descriptive findings on differential mortality. Our study emphasizes the potential of using longitudinal survey samples for differential mortality estimation, in particular for countries where no information is available from other sources or which otherwise have to rely on unlinked cross-sectional data. However, several data quality issues appeared in EU-SILC UDB data for some countries, and some general methodological issues such as proper longitudinal weighting and variance estimation are yet unsolved. Further research on the subject is needed and will be published in two forthcoming reports.

Authors
Johannes Klotz, Statistics Austria, Vienna, Austria (johannes.klotz@statistik.gv.at)
Tobias Göllner, Statistics Austria, Vienna, Austria (tobias.goellner@statistik.gv.at)

Keywords
Mortality; Socio-economic status; EU-SILC; Longitudinal Data; Europe.

1. Background

In an era of pension reforms aiming at increasing the effective retirement age,[1] mortality differences between socio-economic subgroups have become increasingly important. All European countries with available data show that those with higher education, higher incomes and better occupational positions live longer, on average, than the poor. There is also evidence that the substantial mortality improvements during the last decades have not come along with a reduction of socio-economic inequalities in mortality, although there is variation between countries and between absolute vs. relative inequalities (Mackenbach et al. 2016). One may question the fairness of a uniform pension age if the poor are likely to subsidize—via their shorter life expectancy—the pensions of the rich[2] (Bonenkamp 2007). Besides socially undesired redistribution, this has also adverse fiscal consequences (Knell 2016). Then, the sheer feasibility of working longer is doubtful for some population groups among which health deteriorates relatively early over the life course. It is therefore important to analyze not only socio-economic differences in life expectancy, but also in healthy life expectancy (Majer et al. 2011). Apart from its social policy implications, differential mortality by socio-economic status is also important in demographic projections (Jasilionis et al. 2014; van Baal et al. 2016).

A key issue in the analysis of the social gradient of mortality is its variation between countries. Since European countries differ substantially in terms of—for example—pension systems, health systems and labor market policies,[3] a comparative European perspective may allow for important insights into the determinants of social inequalities in mortality. As pointed out by Kulhánová et al. (2014), "cross-country comparisons of socioeconomic

[1] In a more general sense, pension reforms in Europe since the 1990's have essentially aimed at adapting public pension systems to demographic ageing, while at the same time ensuring adequate incomes of the elderly and avoidance of old-age poverty. These targets may be achieved by several measures, not the least because European countries have very different pension schemes. It is clear, however, that in most European countries an increase in the effective retirement age will be a prerequisite for fiscal sustainability of the public pension system during the coming decades (for details see Whiteford & Whitehouse 2006 and Carone et al. 2016). A long-run increase in the effective retirement age may also be seen as fair between generations, given the long-run increase of life expectancy.

[2] Of course, whether a population subgroup are net contributors or net beneficiaries in the public pension system depends on many more parameters than just life expectancy—for instance labor force participation rate, unemployment rate, or mean age at entry into the labor market. It also depends on the construction of the public pension system and may differ between a period and a cohort perspective. Such things are far beyond the scope of this report. What matters here is that ceteris paribus, a uniform pension age disadvantages population groups with lower life expectancy.

[3] Many attempts have been made during the last decades to group European countries by policy regimes, in particular welfare state regimes. The most famous instance is Esping-Andersen (1990).

inequalities in mortality are an important source of information for understanding the mechanisms that generate these inequalities".

A major deficiency in that respect is that mortality by socio-economic status is not part of the European Statistical System (ESS). Whereas all European Union Member States are legally required to transmit harmonized annual data on deaths and population by age and sex to Eurostat, an additional breakdown by educational attainment is voluntary, and other socio-economic breakdowns are not collected by Eurostat at all.[4]

As of 2017, some European countries still cannot provide any figures on socio-economic differences in mortality. Among those which can, figures are not easily comparable, for they differ in terms of data source (unlinked cross-sectional vs. linked longitudinal data), period of analysis, population covered,[5] socio-economic stratification variables, and mortality indicator. Also, since data is not generally produced by National Statistical Institutes (NSIs), communication between countries is not embedded in established ESS structures like Eurostat Working Groups.

Attempts that have been made so far to compare differential mortality between European countries can be classified into three approaches: literature reviews, micro data collection and ex-post harmonization, and macro data collection on mortality by educational level. A detailed literature review is also a deliverable of the FACTAGE project.[6] Literature reviews in general mainly serve as compact overviews of available studies and data sources; however their analytical added value is limited, because the comparability problems mentioned above are rather reproduced than solved.

A more sophisticated approach is micro data collection and ex-post harmonization, as it has been done for two decades by Erasmus MC Rotterdam under various projects funded by EU research funding framework programs led by Prof. Dr. Johan Mackenbach (EUROTHINE, EURO-GBD-SE, DEMETRIQ, and currently LIFEPATH; for exemplary outcomes see Mackenbach et al. 2016). The idea is to first collect country-specific micro data from national data producers (be it NSIs or not). Some pre-specified social stratification variables are then re-coded for each country in such a way to achieve, as best as possible, comparable breakdowns between countries (such as 'high' vs. 'low' educational levels or 'manual' vs. 'non-manual' occupational classes).

Since Erasmus MC Rotterdam has now a decade-long tradition in this field, data is available for a substantial number of countries and also as time series. Moreover, information is also available on the cause of death.[7] The communication channels for data collection and output

[4] Currently (as of 2017), data is collected by Eurostat in the annual UNIDEMO demographic data collection. Deaths and population breakdowns by educational attainment are voluntary tables (D07 and P09).

[5] In some countries, data are available only for certain regions or cities which are not necessarily representative for the entire country. Also, in some studies certain population groups such as non-nationals were excluded.

[6] This literature review is to be produced by the University of the Basque Country; preliminary findings were presented by Amaia Bacigalupe at the FACTAGE expert workshop on differential mortality, which took place from 15 to 17 March 2017 at Statistics Austria in Vienna.

[7] Information has also been collected on other health variables such as self-perceived health, health behaviors and health services use. However, these data come from health interview surveys and are not individually linked with the mortality data.

checking are well-established. However, the approach relies on the availability of national data, so countries which have not produced any figures on the subject on their own are not covered. Also, for some populations unlinked cross-sectional rather than linked longitudinal data have to be used, which are subject to a numerator-denominator bias. As demonstrated by (Shkolnikov et al. 2007; Jasilionis et al. 2012)), the difference in mortality estimates between linked and unlinked sources can be substantial, and the direction of bias when using unlinked sources is in general unknown. And finally, although ex-post harmonization of national classifications works well for many purposes, it is not the same as using harmonized variables in the first place[8], since ex-post harmonization typically means summarizing variables into a few broad categories, which may hide important variation within a category.

Finally, in recent years Eurostat and even more recently the OECD have started macro data collections on mortality rates by educational level (ISCED-97 groups 0-2, 3-4, and 5-6). Precisely, the OECD has collected education-specific mortality rates (Murtin et al. 2017), and Eurostat has collected education-specific deaths counts and related them to either—if available—education-specific population exposures or estimates of them based on Labour Force Survey data (Corsini 2010).

Both data collections now cover a substantial number of countries[9] and use a harmonized educational classification. However, the problem, that for some countries estimates rely on unlinked cross-sectional data, is also present here. Then, relying only on educational level to measure socio-economic position is problematic. Even if the ISCED classification is technically comparable between countries, it may have different social stratification implications[10] (see also Martikainen et al. (2007) for a temporal perspective). Moreover, there is evidence that education is quite valid in terms of behavioral and cultural factors explaining mortality inequalities, but inequalities in material resources are better measured by income. Using only ISCED levels one cannot consider the effect of other social stratification variables, and the comparison is restricted to life expectancy (no healthy life expectancy). Finally, the Eurostat figures on inequalities in life expectancy are calculated on quite strong assumptions. Namely, it is assumed that there are no mortality differences beyond 75 years of age,[11] and that observations (population and deaths) with missing information on educational level are missing at random, so that they can be proportionally allocated.

---

[8] Insofar as countries do not already deliver harmonized variables like ISCED educational groups.

[9] In the Eurostat database, figures for 19 different countries and ISCED-97 groups 0-2, 3-4 and 5-6 are currently available, covering in the best case all single calendar years from 2007 to 2015 (demo_mlexpecedu).

[10] One may think of the generally lower levels of educational attainment in Southern European populations, and of outdated qualifications that were obtained specifically in the context of Socialist economies in Eastern European countries. Also, technical comparability is not fully achieved, given the problems in classifying vocational education in German-speaking countries.

[11] This lies in the fact that for several countries the educational level distribution of the population is estimated from Labour Force Survey data, which in general is restricted to respondents below 75 years of age. So in general with this data no empirical estimates are available on mortality rates by educational level for those aged 75 and over. Assuming equal death rates between educational groups above the age of 75 results in underestimation of life expectancy differences. An alternative approach to assuming equal death rates would be to extrapolate observed mortality differences to higher ages.

Using harmonized EU-SILC longitudinal data?

Our idea is to estimate socio-economic mortality differentials in a comparative European perspective by using harmonized sample data. Precisely, we aim to use the longitudinal component of the European Union Statistics on Income and Living Conditions (EU-SILC) survey, which is the major data source on comparative European statistics on household income and poverty. EU-SILC is embedded in the European Statistical System; it is conducted annually in all European Union Member States (plus some additional countries like Norway) based on Regulation (EC) No 1177/2003 of the European Parliament and of the Council of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC). Most countries have started data collection in 2004, some even in 2003. EU-SILC is an integrated cross-sectional and longitudinal survey. In most countries, households are interviewed in four consecutive calendar years, meaning that each year a quarter of the sample households are replaced by fresh households. Quality assurance is standardized and generally high. All countries are required to fulfil minimum effective sample sizes and deliver harmonized target variables, though the mode of data collection is up to the countries.[12]

The longitudinal component of EU-SILC contains information on vital status (survived or died since the previous year). Though this information is primarily collected for the correct calculation of longitudinal response rates (for those who have died are no longer part of the target population), it can clearly be used for analytical purposes too. EU-SILC's target variables contain, inter alia, information on household income, educational level, and occupation. Moreover, variables on household size and family structure are available, as is information on the dwelling and some geographical breakdown (European Commission - Eurostat 2014b). EU-SILC also contains a 'Minimum European Health Module' (Robine et al. 2010) with items on self-rated health, chronic diseases, and health-related limitations in daily activities. Thus, one may estimate multivariate models, and one may estimate life expectancy and healthy life expectancy by using the same data source. Since survey respondents are interviewed several times, one can also incorporate time-dependent covariates for better understanding of causal mechanisms (Huisman et al. 2005).

So we can see that there would be many advantages in using EU-SILC longitudinal data for comparative European estimation of differential mortality. It would generally increase comparability and provide a minimum standard for those countries which at the moment cannot provide any statistics at all (admittedly, it will hardly add value to national figures for the Nordic countries). But, before we proceed further, two topics have to be dealt with. Firstly, what special issues have to be considered when estimating differential mortality from longitudinal sample data? Secondly, what about the quality of EU-SILC longitudinal data in terms of data accessibility and accuracy of mortality information?

[12] EU-SILC's predecessor, the European Community Household Panel (ECHP), was conducted as a pure longitudinal survey (panel study) in 1994-2001 with a blueprint questionnaire mandatory for all countries. In EU-SILC, countries are allowed greater flexibility on how to obtain data. This is especially important since available data sources vary widely between European countries and also change over time.

2. Estimating differential mortality from sample data

The idea of estimating differential mortality from a sample survey follow-up is not new. In the same fashion as sample surveys are used as a surrogate for complete enumerations in cross-sectional applications (e.g., estimating the size of the labor force from labor force survey data), one may use a sample follow-up as a surrogate for a census follow-up or register data. Estimating differential mortality from longitudinal sample data is also common in clinical studies, although there one hardly works with probability samples.

Several authors estimated differential mortality in Germany based on sample survey data. Lampert et al. (2007) used the German Socio-Economic Panel and applied a fully parametric Gompertz model combined with national life tables. The same methodology was applied to the German National Health Interview and Examination Survey by Hoebel et al. (to be published). Luy et al. (2015) developed a new nonparametric method and applied it to the German Life Expectancy Survey. Then, a feasibility study for Belgium was conducted by Charafeddine et al. (2014), who applied standard life table methods to Health Interview Survey and EU-SILC data. Kulhánová et al. (2014) applied a semiparametric Cox proportional hazards model to data from pooled Dutch Labour Force Surveys.

In most of the examples mentioned above, information on survey participants' vital status[13] was obtained via linkage (be it deterministic or stochastic) with national mortality registers. But vital status information may also be a genuine part of a (longitudinal) survey. Such information is typically recorded to exclude deceased respondents from the calculation of longitudinal response rates, but can of course also be used for analytical purposes. An example is Majer et al. (2011), who used EU-SILC's predecessor, the European Community Household Panel. Their work is particularly interesting for FACTAGE because it aims at internationally comparable figures and uses mortality and morbidity information in an integrated fashion. Mortality information is also recorded in the Survey of Health, Ageing and Retirement in Europe. A first exploration of the potential of this data for differential mortality analyses is Boháček et al. (2015).

A special class of sample follow-up data is when a census follow-up is restricted to certain people enumerated in the census. Examples are the Longitudinal Study in England and Wales[14] and the Permanent Demographic Sample in France (Couet 2007), both linking around 1 percent[15] of all census records with register data on mortality and many other events. Such restricted census follow-ups come with the advantage that they are not subject to genuine sample survey issues like nonresponse.

[13] In theory, one tells those who died from those who have survived until the end of the follow-up period. In practice, among the respondents for which no death is recorded there are not only survivors, but also people who have out-migrated or otherwise been lost to follow-up. It therefore makes sense to distinguish between deaths and *censored* observations.

[14] Detailed information can be found at http://www.ucl.ac.uk/celsius .

[15] The selection refers to four birthdays each.

On the pros and cons of sample follow-up data[16]

Compared to census follow-up or register data, survey follow-up data comes with both pros and cons in differential mortality estimation. On the plus side, samples contain many variables that may be used in socio-economic status operationalization or as control variables. In the case of EU-SILC, one may think of health information or the household context. Socio-economic classifications can be more accurate in sample than census data because of higher quality standards in data collection, more thorough data checking and imputation of missing values. Then, international comparability is usually easier to achieve in the case of survey samples. Finally, for researchers it is much easier to get access to sample data than to census data, where confidentiality is often required by law.

The most obvious drawback of sample data is reduced precision due to smaller death counts. Precision can be somewhat, but not fundamentally, improved by special estimation techniques like model-assisted estimation. Then, there are particular sample data issues such as standard error estimation or proper (longitudinal) weighting. Additionally, mortality estimates from sample follow-up data come with validity issues. It is too often simply assumed that non-coverage of the institutionalized population is the problem, but the situation is more complicated. We must distinguish between validity of general mortality and validity of differential mortality. The former is more a short-run, the latter more a long-run issue.

In a survey which covers only the population in private households and in which participation is voluntary, it is most likely that the survey population is statistically healthier than the general population, insofar as selection by health status is not corrected by sample weights. Nevertheless, a health advantage *at the time of the survey* does not necessarily translate into a mortality advantage in a follow-up period: People who lived in private households in good health at the time of the interview may become ill and/or institutionalized afterwards and then die. The longer the follow-up, the less important health selection into the survey becomes, for its effect on mortality diminishes over time. Nevertheless, one may safely assume that on average, general mortality rates observed in a survey follow-up population are lower than in the general population. This was confirmed by Lampert et al. (2007) for the German Socio-Economic Panel and by Kulhánová et al. (2014) for Dutch Labour Force Surveys.

A downward bias in general mortality does not, however, necessarily imply a bias in differential mortality, at least when measured on a relative scale. If, for example, the true mortality rate of a poor population is 10 and that of a rich population is 5, and in the sample follow-up population both mortality rates are downwardly biased by, say, 20 percent, then the estimated mortality rate ratio is valid (8/4=10/5).[17] Applied to our approach, the question is thus whether the *degree of underrepresentation* of deaths in the sample population differs between the subgroups of interest. Very little is known about this so far. At best, one may get some idea by looking at information such as variation in institutionalization rates or nonresponse rates between subgroups.

[16] This subsection was published almost identically by the same authors in the report "Excess Mortality in the Europe 2020 AROPE Target Group. D1: Methodological Guidelines. NetSILC 3 WP 2.8 Research Paper, Draft Version, May 2017".
[17] The absolute difference in rates would be underestimated, though. For a discussion on absolute vs. relative measurement of health inequalities see Harper et al. (2010).

What *is* known, however, is that socio-economic position is time-dependent.[18] Respondents that were poor at the time of the interview are not necessarily so until they die. Thus, in the long run, the socio-economic stratification at the time of the survey loses relevance, and differential mortality is most likely to be underestimated. The principal fact that social positions may change during the follow-up period is of course not a genuine sample follow-up problem, but affects also census follow-up data. It is, however, accentuated in the sample case, for there the length of the follow-up period is typically chosen longer, in order to guarantee a sufficiently large death count.

3. EU-SILC longitudinal data

Most countries have implemented the longitudinal component of EU-SILC by a four-year panel survey, meaning that respondents are interviewed in four consecutive calendar years and then withdrawn from the sample.[19] The longitudinal component of EU-SILC contains four target variables which are relevant for mortality analyses. DB110 is a household register variable indicating (inter alia) if an entire household has died since the previous survey wave and it thus no longer to be interviewed.[20] RB110 is a personal register variable indicating (inter alia) if in a household which is to be re-interviewed a single person has died, so that for this individual the survey is discontinued. In case of death of a single person, target variables RB140 and RB150 contain month and year of death.

So in principle for each person to be re-interviewed in EU-SILC, one can tell whether this person has survived or died since the previous survey wave. Combining this information on vital status with other target variables in principle allows for estimation of differential mortality parameters, e.g. mortality rates by household income.

User Database (UDB) restrictions

As a researcher, one does not get access to the original EU-SILC data collected by Eurostat, but to the User Database (UDB). The UDB of EU-SILC is enacted by the Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009. This regulation and further information can be found at http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.

Compared with the original data, the UDB comes with limitations. Most importantly, Germany is not included in the UDB. Then, for some target variables, values are grouped or censored. The idea behind these restrictions is to anonymize EU-SILC data of every country for the freely accessible UDB. The anonymization rules are accessible in the document for the longitudinal data (European Commission - Eurostat 2014a). In the following we summarize the restrictions that are relevant for our study.

Restrictions applied to all countries

The UDB adds a variable called RX010 ‘age at time of interview’, which is used in our analysis (alongside our own age at time of interview variable, which we computed by

[18] An advantage of educational level in this respect is that it usually stable in adult life.
[19] Norway has an 8 year follow-up, France 9 years and Luxembourg a permanent panel.
[20] Usually, ‘entire household died’ occurs for single-person households.

subtracting the date of interview from the date of birth).The variables containing information on the year of birth (RB080, PB140) or age at time of interview (RX010) are censored so that everyone who is 80 years old or older belongs to the same age group.

Variables containing the information on the month of an event are grouped into quarters. Relevant for us is the grouping of month of birth (RB070, PB130), household interview (HB050), personal interview (PB100) and when the person moved or died (RB140, except in Belgium, where months are not grouped into quarters).

Restrictions applied to specific countries

The variable month of birth (RB070, PB130) is not provided for Ireland (IE), Malta (MT), the Netherlands (NL), Slovenia (SI) and the United Kingdom (UK). Additionally, the Netherlands (NL) do not provide the month when a person moved or died (RB140).

Malta groups the variables year of birth (RB080, PB140) and age at time of interview (RX010) into five-year bands. A random perturbation of year of birth (RB080, PB140) to group single years of age into age classes (the difference to the true value not exceeding 5 years), and appropriate modifications of related age variables are applied in Finland and Iceland (only some households). This rule also affects age at time of interview (RX010).

In the United Kingdom all records (at household and individual level) pertaining to households of size 10 and over are suppressed.

Validity of mortality information on the micro level (vital status)

As indicated above, it is up to the countries how they get information on the EU-SILC target variables. It is therefore not clear from the outset how mortality information is obtained in each country. Although almost all European countries have a national mortality register, this is not necessarily the (only) data source used.[21] And since mortality measurement is not among the intended uses of EU-SILC data, one may also expect comparably large cross-country variation in the quality of information on respondents' mortality. To get an idea of the quality of vital status information on the micro level, we conducted a small survey among the members of the Eurostat Working Group 'Living Conditions'.

The questionnaire was sent out by the Austrian member of the Working Group by e-mail on 26 January 2017, together with a cover letter. It could be answered either as PDF form with automatic e-mail reply, or in plain text with manual reply. The PDF form is included in the Annex of this report. A reminder mail was sent out to five countries on 15 February 2017. Finally, we got back 35 questionnaires, 28 of them referring to UDB countries[22].

The questionnaire contained five questions. Here we focus on two of them, namely 'How is the information on a respondent's death obtained? (Multiple answers possible)' and 'In your country, is it theoretically possible to link EU-SILC microdata with mortality information

[21] One has to be aware that the decision whether a household or person is eligible for re-interviewing, usually, has to be made during fieldwork. Even if a national mortality register exists in a country, the information available to National Statistical Institutes will hardly be updated daily. So it is almost inevitable that *in the first place*, some mortality information is obtained by the interviewer. What may happen is that this information is later updated or enriched by ex post linkage of EU-SILC data with national mortality register information.

[22] See Table 1 for a list of countries.

from national death registers (for example, via a unique personal ID variable available in both datasets)?'

On the first question, we see that the most important source for mortality information is other household members, which was mentioned by 28 out of the 35 answering countries. 16 countries mentioned the interviewer[23] and 14 linkages with external data sources (mostly population registers). Of minor importance are responses to written sources (response to a letter/notification), neighbors and local authorities. Most countries (22) use several data sources to obtain vital status information. See also Table 2.

On the second question, the most interesting finding is that a majority of all countries are in the position to link EU-SILC microdata with national mortality registers, but only a minority of them has already done it. Precisely, 8 countries mentioned that linkage has already been done, and a further 15 responded that it could be done. Thus, it would be of great importance to encourage those 15 countries to do so. Only 11 countries said that linkage with national mortality register is not possible.[24] It is also interesting that no country mentioned legal problems with linkage, which is often an issue when working with census data. See also Table 3.

Validity of aggregate mortality information on the macro level (death rates)

A third quality issue is validity of general mortality levels in EU-SILC. Even if vital status information is accurate on the micro level, this does not guarantee that aggregate mortality rates are unbiased. The question is whether the sample population is representative, in terms of mortality risk, for the national population.[25]

A useful measure of validity of general mortality levels is the relative mortality ratio, dividing the observed number deaths in the sample follow-up by the expected number of deaths. The expected number is calculated by applying population mortality rates (from available national life tables) to the population at risk of dying in the sample follow-up.

We computed relative mortality ratios applying an existing SAS code provided by Prof. Paul Dickman (see http://biostat3.net/download/sas/relative_survival_using_sas.pdf for the user guide and www.pauldickman.com/rsmodel/sas_colon.zip for the code itself). The default weighting method of individual lifetimes in Dickman's code is Ederer II (Seppä 2014). The population life tables for all countries were extracted from the Eurostat database freely accessible on the internet (demo_mlifetable). We selected single-year survival probabilities (PROBSURV), which are available up to 84 years of age, in the calendar years 2004 to 2014.

[23] We did not ask *how* the interviewer may obtain information on a respondent's death. It is quite likely that this often happens by asking other household members (in which case this category would be redundant), but in general an interviewer may obtain vital status information also from other sources, e.g. from neighbors or from new tenants or house owners.

[24] One country did not answer this question.

[25] What is meant here is the *actual* mortality risk of the sample population, not the measured mortality risk which is, as discussed above, also subject to measurement error. Of course, in our application one cannot tell whether any deviation of an observed from an expected death count is caused by measurement error on the micro level or non-representativity of the sample on a macro level. Nevertheless, it is important to understand the difference between the two.

Three caveats of this method have to be considered. Firstly, the available algorithm assumes that a person aged 77 at the time of the interview is exactly 77 years old (i.e., that the interview day is his or her birthday) and not somewhere between 77 and 78 years. This means that on average we slightly underestimate the individual probability of dying and hence, the expected number of deaths. Secondly, the available algorithm does not allow for weighting observations with sample weights, so conditional on calendar year, sex and age group, we use unweighted data. What effect this has on the expected number of deaths is unclear. Thirdly, the Ederer II weighting is also debatable, see Seppä (2014). For the final report, we aim to develop an improved algorithm. For the time being, we believe that the relative mortality ratios output by the algorithm should allow for a sufficiently sound quality rating at least of the order of magnitude of aggregate (general) mortality figures.

A relative mortality ratio of 1.0 means that the survey population is as mortal as the general population (conditional on sex, age and calendar year). A ratio greater than 1.0 means that more deaths were observed in the sample follow-up than could be expected based on population mortality rates. We focus on cumulative mortality rates after three years of follow-up (which in most countries is the maximum follow-up period).

4. Data preparation

In this section we describe in detail how the data available in the EU-SILC UDB was prepared for mortality analysis. A graphical overview is given in Figure 1 and Figure 2. We used SAS, Version 9.4. SAS code is available from the authors on request.

In a longitudinal survey where respondents are interviewed several (usually four) times, different data sets can be constructed, depending on the statistical model one wants to apply. Here we decided on analyzing individual lifetimes, i.e. each record in the data file corresponds to a distinct person, and the maximum observation time for this person from the first interview to either death or censorship is recorded. This is the most common kind of data in longitudinal mortality analyses.[26]

Longitudinal datasets in the UDB are labeled with the latest of four calendar years in which respondents belong to the sample, i.e. the longitudinal dataset 2006 covers respondents that were first interviewed in 2003 and then to be re-interviewed in 2004, 2005 and 2006. Households or persons who were withdrawn before the latest year, most importantly those who died between two interviews, are also included in the data. For our analysis we extracted longitudinal datasets from 2006 until 2014, meaning that our period of analysis ranges from 2003 to 2014. Datasets in the UDB are occasionally updated when new information becomes available. We used the most up-to-date versions that were available to us at the time of query in January 2017. For details see Table 4.

[26] An alternative would be to prepare the data such that each record corresponds to a person and a year of the survey. The advantage for differential mortality analyses is that one could then also incorporate individual changes in social status between successive survey waves. However, observations would then no longer be statistically independent, so this would be more difficult to model. See for instance Majer et al. (2011).

For quality assurance the following documents were scanned: 'SILC Longitudinal UDB Problems and Modifications' (2006-2014); 'SILC_ESQRS' documents, these are the annual comparative quality reports of Eurostat (2006-2014); and 'Differences between Data Collected and UDB' (2009, 2013 and 2014).

Our analysis is based on individual lifetimes, so we need to know certain date values for each person in the sample (in particular, date of first interview and date of death or censorship). Since the UDB contains only years and quarters, we imputed the midpoint of each quarter. For instance, for a date in the second quarter we assume 15 May.[27]

In the UDB Greece has two different country codes, 'GR' in the datasets 2006-2008 and 'EL' in the datasets 2009-2014. We merged both codes into 'GR'.

EU-SILC data is split into four files: D, H, P and R. The D-file is the household register, which contains every household drawn into the sample even if the household could not be contacted or interviewed. The H-file is the household data, which contains every household that has been successfully contacted and interviewed. The R-file is the personal register, which contains every person currently living in the household or temporarily absent. The P-file is the personal data, where every person with a personal interview is stored. Since we are interested in the deaths of people, we use the R-file as a base file and enrich it with information from the D-, H-, and P-file. This was done by deterministic linkage using as linkage variables country, year of survey and ID of household or person.[28]

A total of 1,175,718 distinct individuals with at least one longitudinal observation are available in our data. After some exclusion criteria (see below), we are finally left with 1,039,611 distinct persons. The total of person-years observed is 2.4 million.

Deaths can be identified by either DB110=5 (if an entire household dies) or RB110=6 (if a person dies, but at least one other household member survives). Double counting is in principle excluded, since if an entire household moves or dies, then no person record should be created any more. There were however 95 instances where the D-file indicates movement of the entire household, while the R-file indicates death of a person. We assumed that the D-file is the first one to be filled and although the death event may be correct we did not count it in this case. Our data contains 24,002 deaths of distinct people. In the case of DB110=5 we do not know the month (quarter) and year of death, so we assumed the midpoint between the date of last interview and the date to be re-interviewed.

However, not all of the 24,002 death events are useable. Figure 3 depicts how many cases per country we had to omit because of the following exclusion criteria: We had to exclude

[27] An alternative would be to impute a random date in the quarter. Then, however, one has to implement additional consistency rules, for example that a person does not die before he or she is interviewed.

[28] The ID of a person is the ID of the household plus two additional digits identifying a person within a household. The household IDs should be distinct within countries, but not necessarily between them. We checked the R-file and noted that there are some countries that reassign once used IDs again. See Table 5 for a list of all countries and their respective numbers of cases where this happened.

persons which are 80 or older at baseline,[29] since in the UDB they are grouped into one age category. This reduces our death count by almost a third to 16,276.[30] We further excluded persons under 16 years of age, because for most of them no personal interview (containing e.g., health information) is available. Finally, we excluded cases where the 'time at risk' variable, which is calculated as the duration between first interview and the date of death or censorship, exceeds the maximum plausible value of usually 4 years (for exceptions see Table 6). This mainly concerns the cases where the personal ID was reassigned to a new respondent, resulting in a loss of about 8 percent of all deaths. The final number of deaths is 14,711. For country-specific figures see Table 6.

5. Findings on feasibility

Death counts

A cross tabulation of country and year of death is given in Table 7. We see that for some countries no deaths are recorded in some calendar years. This is trivially the case for the years before the survey was implemented in this country, for example in Croatia before 2010. In the same fashion, some countries were not (yet) included in the first version of the 2014 longitudinal dataset, explaining for instance the zero death count of Croatia in 2014. In some other cases however, mortality information is discontinued, meaning that records on respondents who died were not transmitted to the UDB at all. This was the case in Denmark, Ireland and Iceland after 2008.

Since 2008, the annual total of usable deaths has ranged between 1,600 and 1,900. The lower count of 900 in 2014 is mainly due to the fact that interviews are distributed over the year, and deaths in the last year are included only when they occurred *before* the re-interview date. Comparing countries, more than 1,000 deaths are in total available for Poland, Italy, Hungary and Slovenia. As a rule of thumb, 500 deaths should allow for sufficiently reliable figures on differential mortality. This is further possible in Spain, France, Czech Republic, Latvia, Bulgaria, Finland, Estonia, Greece and Lithuania (see also Figure 3).

Relative mortality figures

Figure 4 and Table 8 depict the cumulative relative mortality ratios after three years of follow-up. Ireland, Iceland and Denmark were excluded from the calculations due to discontinuity of mortality information in the UDB data. Out of the remaining 26 countries,

---

[29] Age was measured by the UDB variable RX010 'Age at the time of interview'. If this variable was missing, then we imputed the integer part of the difference of the date of interview minus the date of birth (both dates given in months and years).

[30] Clearly, the grouping of ages 80 and over into one category has greater impact on mortality analyses than on most other uses of EU-SILC data. Note, however, that although the proportion excluded is very high regarding the death count, it is much smaller regarding the potential years of life lost. Also, for the highest age groups survey errors (nonresponse, non-coverage of the population in institutions) are typically large, and established techniques are available to extrapolate death rates beyond the age of 80, so the impact of this exclusion on differential mortality estimates will be smaller than one may assume at first glance. In future analyses of our data, we will restrict, from the very start, the sample to those aged 16-79 at baseline. However since the subject of this report is feasibility of mortality estimation, we decided to start here with all deaths. Some researchers might be interested in mortality differences among the oldest-old, and it should be clarified here that UDB data is not usable for that.

most show relative mortality figures below 1.0, meaning that the sample population is less mortal than the general population. In principle, this is quite likely, given that EU-SILC is restricted to the population in private households (at the time of interview) and that bad health (e.g., being hospitalized at the interview day) increases the probability of survey nonresponse.

There are, however, some countries with values far below 1.0, indicating that there may be not only a survey bias, but also a quality problem in mortality information or some restrictions in UDB information which we do not fully understand yet. Less than half of the expected death count is observed for Norway, Belgium, the United Kingdom, and in particular for males also for the Netherlands and Romania. This list of countries is rather surprising, since for most of them data quality can—in general—be assumed to be high.

There are also some countries where the sample population shows a higher death count than expected, more often for females than males. The greatest deviation is observed, for both sexes, for Slovenia. It should be noted that Slovenia uses a special sample for EU-SILC (person sample instead of household sample) and this may influence results in a way that we do not fully understand yet.

For 21 out of the 26 countries, we observe a higher relative mortality ratio among the female than the male population. The unweighted average is 0.80 for males and 0.99 for females. It may be that sex modifies the correlation of mortality risk with nonresponse.[31] It may also be that the effect of excluding the population outside private households has a stronger impact on relative mortality ratios among males. Finally, note that here we have used unweighted data and that findings may somewhat change when using weights, as we will do for the final report.

Age-specific mortality patterns

Table 9 shows age-specific relative frequencies of deaths,[32] both total and broken down by sex. Plotting the relative frequency of deaths against the age at first survey, we get an impression of the age profile of mortality rates in our sample data. Figure 5 contains such plots for the pooled data set containing all 29 countries. From age (at first survey) 35 onwards we see the familiar exponential increase of mortality risk with age. This finding can be used in model specification and for extrapolation of mortality rates for higher ages. We also see that male mortality exceeds female mortality in all age groups.

[31] The findings of Bopp et al. (2014) on those who could not be contacted make aware of this.

[32] As mentioned above, our data is prepared so that each record refers to a distinct person who has either survived, died, or out-migrated at the end of the observational period.

6. Some exemplary descriptive findings on differential mortality

Finally, we present some exemplary descriptive findings on differential mortality estimated from our data. The idea is to point at the rich analytical possibilities that our data offers. We do not put great emphasis on social stratification or model selection, since such issues will be part of the next report. Also for now we refrain from longitudinal weighting and variance estimation.

Mortality by income quintile in Poland and Bulgaria

We estimated Cox proportional hazards regression models (Kleinbaum & Klein 2010) for two countries, Poland and Bulgaria. These countries were chosen because death counts and relative mortality figures indicate good data quality, and the scientific literature on differential mortality in these countries is scarce.

The dependent variable is the time between first interview and the end of individual follow-up. Death was modeled as the events of interest, whereas migration and survival were modeled as censoring events. Explanatory variables are sex, age and income quintile.[33] The analysis was restricted to people aged 35-79 at the time of first interview. All available survey years were used in the analysis, so the reference period is 2006-2014 for Bulgaria and 2005-2014 for Poland.

Mortality hazard ratios by income quintile[34] are given in Figure 6. In Bulgaria, people in the lowest income quintile have a mortality risk 1.53 as high as people in the highest income quintile. The relative difference is even larger in Poland, where mortality risk in the lowest income group is 1.76 times as high as in the highest one.

Income or health – what matters for Polish males?

As mentioned above, one of the advantages of the EU-SILC survey is its many variables that are relevant for differential mortality and morbidity analyses. For instance, EU-SILC contains also a 'Minimum European Health Module' (Robine et al. 2010) with three questions on health. One of these questions is on general activity limitations in activities people usually do because of health problems for at least the past six months ('GALI', target variable PH030). This variable is often used in international comparisons on healthy life expectancy, for instance by the REVES network (http://reves.site.ined.fr/en/home).

An interesting question is to what extent health status is correlated with low income, and how this affects mortality risk. A very simple model is to categorize people according to a fourfold table on income (below/above median income) and GALI [35](no/yes). We did so for Polish males, and applied then the same regression model class as above.

[33] 'Income' means equivalized disposable household income (HX090). The five groups were built by applying the RANK procedure in SAS to HX090 in the UDB.

[34] We do not present hazard ratios by age and sex, for here these are only control variables. As expected, mortality increases with age and is higher for males than for females.

[35] The target variable PH030 contains three categories. For our purposes we pooled 'Yes, strongly limited' and 'Yes, limited', to contrast it with 'No, not limited'.

Figure 7 indicates that both low income and activity limitations influence mortality risk, however, those with low income and no activity limitation are better off than those with high income and activity limitation. This is comparable to findings by Huisman et al. (2005) that "higher education serves to postpone or avoid disability, but provides less benefit when disability is already present."

Excess mortality of low educated males by country

A higher mortality risk of people with low education (ISCED 0-2) compared to people with high education (ISCED 5-6) is well documented (Corsini 2010; Murtin et al. 2017) and has been found basically for all countries with available data. The question is now whether our data fit this, too. For that purpose we estimated hazard ratios in the same manner as above. The results we present are restricted to males. Results are given in Figure 8.

Estimated hazard ratios exceed 1.0 in all 26 countries and 2.0 in 14 countries. The highest estimates are obtained for Slovakia and Luxembourg, the lowest estimates for the United Kingdom and Sweden. In general, excess mortality is high in Central Europe and in the Baltic countries. This partly resembles what is known from the literature, although some findings (small inequalities in Portugal and Romania, large inequalities in Luxembourg) seem implausible and could result not only from small death counts, but also from data validity issues.

7. Conclusions and outlook

In an era of pension reforms aiming at increasing the retirement age, mortality differences between socio-economic groups become increasingly important. Statistics on this are, however, not part of the European Statistical System. Although considerable progress has been made on international coverage and comparability during the last years, there are still some European countries which cannot provide any figures on the subject, and available figures are often not comparable (e.g., linked longitudinal vs. unlinked cross-sectional data).

A possible solution to overcome these comparability issues is to use longitudinal sample data with harmonized target variables. The EU-SILC survey has been conducted annually for around ten years now in all EU member states plus some additional countries. It contains harmonized target variables based on European regulation, many of them relevant for differential mortality. It has also a longitudinal component (mostly four-year panel) which records if a respondent dies between two survey waves. So in theory one should be able to use EU-SILC data to achieve comparable statistics on differential mortality between European countries, apart from sampling error.

An important caveat of our approach is that researchers do not get access to the original EU-SILC data transmitted to Eurostat by National Statistical Institutes, but to a special User Database. This User Database comes with certain restrictions in coverage and content, most notably the absence of data for Germany and the grouping of people aged 80 and over into one age group. Some countries are in principle included in the User Database, but mortality information is discontinued there.

As to the quality of mortality information on the micro level, the major data source is other household members. A majority of countries would technically be able to link EU-SILC microdata ex post with national mortality registers, but so far only a minority of them actually did this.

On the macro level, deaths are on average underrepresented by 20 percent among males and by 1 percent among females. This can to some degree be expected due to non-coverage of the institutionalized population and health-related nonresponse; but relative mortality figures also suggest a serious general underrepresentation of deaths in some countries, which we do not fully understand yet. However, a bias in general mortality does not necessarily imply a bias in differential mortality, at least when it is measured on a relative scale.

Extracting the data from the User Database and editing it so that it can be used for mortality analyses was quite complex, but eventually we achieved an algorithm that satisfies our purposes. In particular, one may stratify the EU-SILC sample by any target variable, allowing also for multivariate models. At present there is a restriction in place, namely that time-dependent covariates such as income or health status are taken only from the first interview of a household (the 'baseline', as it is called in epidemiological follow-up studies). A natural extension which we want to implement in the near future is to edit also time variation in covariates. This will be particularly helpful in specifying models aiming to detect cause-effect relationships.

In a sense, EU-SILC data is restricted in that it contains only information on all-cause mortality, but not on specific causes of death. In analyses we are thus constrained to overall mortality measures such as life expectancy differences. One should, however, bear in mind that cause-specific mortality comparisons between almost 30 countries would suffer not only from small death counts, but also from cultural disparities in cause of death registration and coding, so for the purpose of our analysis we believe this constraint to be of minor importance.

An important task which we have not dealt with so far is proper longitudinal weighting and, related to that, variance estimation of estimators.[36] These issues will be solved in the final report. Variance estimation depends of course on the estimator of mortality rates used. We also plan to compare different estimators, and to combine findings from sample data with national life tables to transform excess mortality into life expectancy difference estimates.[37] We also want to discuss in more detail how the socio-economic status can be defined based on EU-SILC data.

The purpose of this feasibility study was to check the potential of EU-SILC longitudinal data for comparative European analyses on differential mortality. The main advantages of EU-SILC longitudinal data, compared to existing alternative data sources, are its universal coverage of EU member states (albeit Germany is not included in the User Database) and the

[36] Treating the data as if it were a simple random sample is not recommended, for this may bring substantial distortions in standard errors (see National Center for Health Statistics (2013) for a comparable application).
[37] An advantage of analyzing differences in years of (healthy) life expectancy is the clarity of this metric. So one may also circumvent the discussion of absolute vs. relative inequalities in health (Harper et al. 2010).

large number of relevant and harmonized target variables. However, several data quality problems have been detected in User Database data for some countries, and some general methodological issues such as proper longitudinal weighting and variance estimation are yet unsolved. Thus, further research is needed to assess whether comparative European estimates on differential mortality can actually be achieved. Nevertheless, our method clearly allows for one thing, namely it may serve as a minimum standard for those countries which currently do not provide any national figures on mortality differentials by socio-economic status. It may also be an important complement for countries where available figures rely on unlinked cross-sectional data.

Outlook

This report is the first in a series of three reports to be produced in FACTAGE Work Package 4. A second report will follow in early 2018, with a focus on results and implications for public pension systems. Finally, a detailed technical report will be published in mid-2018, handling in detail the methodical problems that have so far been left out (in particular, weighting and standard error calculation).

Statistics Austria will host a FACTAGE training session on 25 and 26 April 2018 in Vienna. Representatives of National Statistical Institutes (also from EU candidate countries), other bodies of official statistics (e.g., central banks) and researchers with a connection to official statistics will be invited to learn about the method we developed for the FACTAGE project, so they can apply it to their own data. We hope that this will contribute to a development to eliminate 'white spots' on the European map on differential mortality.

This report will be sent to all members of the Eurostat Working Group 'Living Conditions', first to thank them for their collaboration in the web survey, and then also to provide the possibility to comment on the findings for their country or on the report in general. We also aim to bring up some issues found in our work within the framework of the European Statistical System—in particular, to encourage all countries where this is possible to link EU-SILC micro data with mortality registers to obtain accurate vital status information. Although register information may not yet be available at fieldwork time, it should in most cases be available when data are transmitted to the Eurostat.

It has recently been proposed to delete the target variables RB140 and RB150 (month and year when a sample person moved out or died) from the EU-SILC nucleus. Statistics Austria has already responded to this proposal, pointing out that these target variables are essential for any analysis of mortality and their deletion would substantially reduce the information content for such analyses. In the medium run, we also want to bring up for discussion whether some restrictions of the User Database (such as the grouping of all respondents aged 80 and over into one category) are actually necessary or could be relaxed somehow.

Publication bibliography

Boháček, Radim; Crespo, Laura; Mira, Pedro; Pijoan-Mas, Josep (2015): The educational gradient in life expectancy in Europe: preliminary evidence from SHARE. In *SHARE Working Paper Series* (22).

Bonenkamp, Jan (2007): Measuring lifetime redistribution in Dutch occupational pensions. The Hague: CPB Netherlands Bureau for Economic Policy Analysis (CPB discussion paper, no 81).

Bopp, M.; Braun, J.; Faeh, D.; Egger, M.; Spoerri, A.; Zwahlen, M. et al. (2014): Variation in Mortality Patterns Among the General Population, Study Participants, and Different Types of Nonparticipants: Evidence From 25 Years of Follow-up. In *American Journal of Epidemiology* 180 (10), pp. 1028–1035.

Carone, Giuseppe; Eckefeldt, Per; Giamboni, Luigi; Laine, Veli; Pamies Sumner, Stéphanie (2016): Pension reforms in the EU since the early 2000's. Achievements and challenges ahead. Luxembourg: Publications Office of the European Union (European economy / European Commission, Economic and Financial Affairs, 042 (December 2016)). Available online at http://dx.doi.org/10.2765/620267.

Charafeddine, Rana; Berger, Nicolas; Demarest, Stefaan; van Oyen, Herman (2014): Using mortality follow-up of surveys to estimate social inequalities in healthy life years. In *Population Health Metrics* (12).

Corsini, Veronica (2010): Highly educated men and women likely to live longer: Life expectancy by educational attainment. In *eurostat: Statistics in focus* (24).

Couet, Christine (2007): INSEE's Permanent Demographic Sample (EDP). In *Courrier des statistiques, English series* (13).

Esping-Andersen, Gøsta (1990): The Three Worlds of Welfare Capitalism. Princeton, New Jersey: Princeton University Press.

European Commission - Eurostat (2014a): Differences Between Original Database And Anonymised User Database. SILC Disclosure Control Rules.

European Commission - Eurostat (2014b): Methodological Guidelines and Description of EU-SILC Target Variables. DocSILC065 (2014 operation).

Harper, Sam; King, Nicholas B.; Meersman, Stephen C.; Reichman, Marsha E.; Breen, Nancy; Lynch, John (2010): Implicit Value Judgments in the Measurement of Health Inequalities. In *The Milbank Quarterly* (1), pp. 4–29.

Huisman, Martijn; Kunst, Anton; Deeg, Dorly; Grigoletto, Francesco; Nusselder, Wilma; Mackenbach, Johan (2005): Educational inequalities in the prevalence and incidence of disability in Italy and the Netherlands were observed. In *Journal of Clinical Epidemiology* 58 (10), pp. 1058.e1.

Jasilionis, Domantas; Shkolnikov, Vladimir M.; Andreev, Evgueni M.; Jdanov, Dmitri A.; Vågerö, Denny; Meslé, France; Vallin, Jacques (2014): Do Vanguard Populations Pave The

Way Towards Higher Life Expectancy Among Other Population Groups? In *Population* 69 (4), pp. 589.

Jasilionis, Domantas; Stankuniene, Vlada; Ambrozaitiene, Dalia; Jdanov, Dmitri A.; Shkolnikov, Vladimir M. (2012): Ethnic mortality differentials in Lithuania: contradictory evidence from census-linked and unlinked mortality estimates. In *J Epidemiol Community Health* 66 (6), pp. e7.

Kleinbaum, David G.; Klein, Mitchel (2010): Survival analysis. A self-learning text. 2. ed., [Nachdr.]. New York [u.a.]: Springer (Statistics for biology and health).

Knell, Markus (2016): Grundlagen eines soliden und solidarischen Pensionskontensystems. In *Wirtschaft und Gesellschaft* (3), pp. 465–496. Available online at http://wug.akwien.at/WUG_Archiv/2016_42_3/2016_42_3_0465.pdf.

Kulhánová, Ivana; Hoffmann, Rasmus; Eikemo, Terje A.; Menvielle, Gwenn; Mackenbach, Johan P. (2014): Educational inequalities in mortality by cause of death: first national data for the Netherlands. In *Int J Public Health* 59 (5), pp. 687–696.

Lampert, Thomas; Kroll, Lars Eric; Dunkelberg, Annalena (2007): Soziale Ungleichheit der Lebenserwartung in Deutschland. In *Aus Politik und Zeitgeschichte* (42), pp. 11–18.

Luy, Marc; Wegner-Siegmundt, Christian; Wiederman, Angela; Spijker, Jeroen (2015): Life expectancy by education, income and occupation in Germany: Estimations with the Longitudinal Survival Method. In *Comparative Population Studies* 40 (4), pp. 399–436.

Mackenbach, Johan P.; Kulhánová, Ivana; Artnik, Barbara; Bopp, Matthias; Borrell, Carme; Clemens, Tom et al. (2016): Changes in mortality inequalities over two decades: register based study of European countries. In *BMJ*, pp. i1732.

Majer, I. M.; Nusselder, W. J.; Mackenbach, J. P.; Kunst, A. E. (2011): Socioeconomic inequalities in life and health expectancies around official retirement age in 10 Western-European countries. In *Journal of Epidemiology & Community Health* 65 (11), pp. 972–979.

Martikainen, P.; Blomgren, J.; Valkonen, T. (2007): Change in the total and independent effects of education and occupational social class on mortality: analyses of all Finnish men and women in the period 1971-2000. In *Journal of Epidemiology & Community Health* 61 (6), pp. 499–505.

Murtin, Fabrice; Mackenbach, Johan; Jasilionis, Domantas; Mira d'Ercole, Marco (2017): Inequalities in longevity by education in OECD countries: Insights from new OECD estimates. In *OECD Statistics Working Papers* 2017/02.

National Center for Health Statistics (2013): Analytic Guidelines for NCHS 2011 Linked Mortality Files. Available online at http://www.cdc.gov/nchs/data_access/mortality/linkage_methods_analytical_support.htm.

Robine, Jean-Marie; Jagger, Carol; van Oyen, Herman; Cambois, Emmanuelle; Doblhammer, Gabriele; Nusselder, W. J. et al. (2010): The Minimum European Health Module. Background documents. In *EHEMU Technical Report*.

Seppä, Karri (2014): Relative Survival Estimators. Finnish Cancer Registry. Available online at https://ec.europa.eu/jrc/sites/jrcsh/files/epaac-wp9-session3-seppa.pdf, checked on 6/21/2017.

Shkolnikov, Vladimir M.; Jasilionis, Domantas; Andreev, Evgeny M.; Jdanov, Dmitri A.; Stankuniene, Vladislava; Ambrozaitiene, Dalia (2007): Linked versus unlinked estimates of mortality and length of life by education and marital status: Evidence from the first record linkage study in Lithuania. In *Social Science & Medicine* 64 (7), pp. 1392–1406.

van Baal, Pieter; Peters, Frederik; Mackenbach, Johan; Nusselder, Wilma (2016): Forecasting differences in life expectancy by education. In *Population Studies* 70 (2), pp. 201–216.

Whiteford, Peter; Whitehouse, Edward (2006): Pension Challenges and Pension Reforms in Oecd Countries. In *Oxford Review of Economic Policy* 22 (1), pp. 78–94.

Annex

Table 1

Countries that answered the Eurostat Working Group web survey

UDB countries (n=28)	Non UDB countries (n=7)
Austria	Albania
Belgium	Germany
Bulgaria	Kosovo
Croatia	Montenegro
Cyprus	Serbia
Czech Republic	Switzerland
Denmark	Turkey
Estonia	
Finland	
France	
Greece	
Hungary	
Iceland	
Ireland	
Italy	
Latvia	
Lithuania	
Luxembourg	
Malta	
Netherlands	
Norway	
Poland	
Portugal	
Romania	
Slovakia	
Slovenia	
Spain	
United Kingdom	

Source: Statistics Austria

Table 2

Answers to Eurostat Working Group web survey question 1

How is the information on a respondent's death obtained? (Multiple answers possible)	UDB countries:		non UDB countries:	
Answer	**Countries (n)**	**Percent of countries**	**Countries (n)**	**Percent of countries**
Interviewer	14	50.00%	2	28.57%
Other household members	23	82.10%	7	100.00%
Written source (response to a letter/notification)	3	10.70%	1	14.29%
Linkage with external data sources, namely	12	42.90%	2	28.57%
Other, namely	1	3.60%	1	14.29%

Source: Statistics Austria

Table 3

Answers to Eurostat Working Group web survey question 2

In your country, is it theoretically possible to link EU-SILC microdata with mortality information from national death registers (for example, via a unique personal ID variable available in both datasets)?	UDB countries:		non UDB countries:	
Answer	**Countries (n)**	**Percent of countries**	**Countries (n)**	**Percent of countries**
Yes, this has already been done.	8	29.63%	0	0.00%
Yes, but so far it has not been done.	11	40.74%	4	57.14%
Technically possible, but not allowed for legal reasons.	0	0.00%	0	0.00%
No, not possible.	8	29.63%	3	42.86%
Sum	27	100.00%	7	100.00%

Source: Statistics Austria

Table 4

The longitudinal data in the EU-SILC User Database

Year	Version	Date	Data from	to
2006	2	01.03.2009	2003	2006
2007	5	01.08.2011	2004	2007
2008	4	01.03.2012	2005	2008
2009	4	01.03.2013	2006	2009
2010	4 *	01.03.2014	2007	2010
2011	5	01.08.2016	2008	2011
2012	4	01.08.2016	2009	2012
2013	3	01.08.2016	2010	2013
2014	1	01.08.2016	2011	2014

The most recent versions were used where possible. Note, that the data range is not applicable to all countries, since countries implemented EU-SILC at different times (see Table 6). Source: Statistics Austria
* = version 5 should be available; we try to fix this in the future.

Table 5

Number of reassigned personal IDs

Country	Number of Reassigned IDs
Estonia	2,293
Spain	9,129
Finland	2,749
France	25,595
Lithuania	6,828
Luxembourg	8,925
Malta	18
Norway	14,093
Poland	11,086
Portugal	18,201
Romania	1,901
Slovakia	9,282
Sum	110,100

Source: Statistics Austria

Table 6

Meta data information per country

				Initial Data			Useable Data		
		Start Year*	**Maximum Length of the Follow Up (Years)**	**Number of Persons**	**Sum of Person Years**	**Number of Deaths**	**Number of Persons**	**Sum of Person Years**	**Number of Deaths**
UDB									
AT	Austria	2003	4	32,430	72,247	586	30,706	69,118	371
BE	Belgium	2003	4	32,278	71,160	229	30,726	68,275	160
BG	Bulgaria	2006	4	27,790	63,558	1,147	25,954	60,101	743
CY	Cyprus	2005	4	26,578	57,341	358	25,184	54,805	214
CZ	Czech Republic	2005	4	45,794	114,415	1,193	43,063	108,986	808
DK	Denmark	2003	4	30,810	62,307	137	29,664	60,468	109
EE	Estonia	2004	4	33,392	95,482	972	28,884	67,730	621
ES	Spain	2004	4	103,436	251,180	1,937	87,320	182,391	919
FI	Finland	2004	4	56,159	121,150	892	51,840	110,091	716
FR	France	2004	9	47,058	168,199	1,178	44,032	159,947	810
GR	Greece	2003	4	47,965	103,695	1,058	44,557	97,187	571
HR	Croatia	2010 (2009)	4	16,438	24,561	352	15,393	23,177	243
HU	Hungary	2005	4	56,259	126,918	1,531	53,242	121,376	1,058
IE	Ireland	2003	4	15,534	26,014	37	11,308	18,587	15
IS	Iceland	2004	4	8,584	17,578	71	8,152	16,676	58
IT	Italy	2004	4	123,056	269,000	2,545	114,636	253,417	1,269
LT	Lithuania	2005	4	24,421	80,067	1,073	16,528	36,043	546
LU	Luxembourg	2003	12 **	22,111	74,971	379	21,431	73,468	325
LV	Latvia	2005	4	32,269	70,031	1,203	29,988	65,903	797
MT	Malta	2005	4	19,538	45,843	294	18,905	44,633	270
NL	Netherlands	2005	4	51,801	111,715	346	45,444	100,633	259

		Start Year*	Maximum Length of the Follow Up (Years)	Initial Data: Number of Persons	Initial Data: Sum of Person Years	Initial Data: Number of Deaths	Useable Data: Number of Persons	Useable Data: Sum of Person Years	Useable Data: Number of Deaths
NO	Norway	2003	8	28,771	92,038	163	27,309	88,687	127
PL	Poland	2005	4	76,983	249,123	2,184	61,781	145,177	1,323
PT	Portugal	2004	4	27,939	136,727	656	10,279	18,336	154
RO	Romania	2007	4	34,570	87,408	614	30,620	70,023	383
SE	Sweden	2004	4	34,880	78,956	527	32,862	75,222	318
SI	Slovenia	2005	4	46,805	97,809	1,387	44,603	94,029	1,048
SK	Slovakia	2005	4	24,746	95,842	564	14,663	33,821	242
UK	United Kingdom	2005	4	47,323	92,359	389	40,537	80,765	234
	Sum	n/a	n/a	1,175,718	2,957,693	24,002	1,039,611	2,399,071	14,711
non UDB									
DE	Germany	2005	4	n/a	n/a	n/a	n/a	n/a	n/a
TR	Turkey	2006 (2005)	4	n/a	n/a	n/a	n/a	n/a	n/a
CH	Switzerland	2007	4	n/a	n/a	n/a	n/a	n/a	n/a
RS	Serbia	2013	4	n/a	n/a	n/a	n/a	n/a	n/a
--	FYROM	2011 (2010)	4	n/a	n/a	n/a	n/a	n/a	n/a

Source: Statistics Austria

* In parentheses are the years with a trial run

** Luxembourg has a permanent panel, so the maximum number of follow-up years is 12 right now. This will change over the years.

Table 7

Number of useable deaths per year and per country

Country	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Sum
AT	-	14	45	44	58	37	32	28	40	32	26	15	371
BE	-	4	23	15	16	7	10	6	16	39	24	-	160
BG	-	-	-	10	40	60	88	128	141	111	103	62	743
CY	-	-	5	22	25	19	21	23	22	31	24	22	214
CZ	-	-	23	66	95	110	140	93	90	106	85	-	808
DK	1	27	21	23	15	22	-	-	-	-	-	-	109
EE	-	19	40	68	63	53	73	60	66	56	73	50	621
ES	-	16	59	89	70	92	87	148	98	100	108	52	919
FI	-	54	49	58	58	70	35	44	70	97	119	62	716
FR	-	27	81	107	74	79	82	74	87	82	65	52	810
GR	3	10	42	68	65	73	48	56	89	32	40	45	571
HR	-	-	-	-	-	-	-	14	68	84	77	-	243
HU	-	-	20	96	106	148	111	106	139	124	141	67	1,058
IE	-	-	-	1	10	4	-	-	-	-	-	-	15
IS	-	4	15	25	7	7	-	-	-	-	-	-	58
IT	-	31	96	118	128	120	115	169	184	170	99	39	1,269
LT	-	-	13	40	62	62	61	44	49	67	81	67	546
LU	15	20	23	30	34	37	29	19	38	31	25	24	325
LV	-	-	24	67	72	80	90	97	104	102	90	71	797
MT	-	-	-	3	13	35	37	37	45	28	32	40	270
NL	-	-	31	25	37	25	40	20	33	25	23	-	259
NO	1	4	2	10	13	20	28	14	13	11	10	1	127
PL	-	-	21	106	169	162	177	162	163	139	109	115	1,323
PT	-	-	1	10	12	20	5	15	15	20	28	28	154
RO	-	-	-	-	9	59	69	41	80	56	39	30	383
SE	-	-	46	34	37	56	59	48	23	7	8	-	318
SI	-	-	-	-	75	121	151	159	186	173	137	46	1,048
SK	-	-	6	20	15	12	44	19	21	37	68	-	242
UK	-	-	-	13	21	39	42	36	25	15	20	23	234
Sum	20	230	686	1,168	1,399	1,629	1,674	1,660	1,905	1,775	1,654	911	14,711

Source: Statistics Austria

Table 8

Cumulative relative mortality ratios after 3 years of follow-up

Country	Male	Female
Austria	0.90	0.95
Belgium	0.38	0.39
Bulgaria	1.00	1.01
Cyprus	0.78	0.90
Czech Republic	0.82	0.86
Denmark	n/a	n/a
Estonia	0.93	1.02
Spain	0.87	1.12
Finland	0.92	1.61
France	0.75	0.94
Greece	0.79	0.73
Croatia	0.99	0.87
Hungary	0.86	0.84
Ireland	n/a	n/a
Iceland	n/a	n/a
Italy	0.80	1.04
Lithuania	1.02	1.21
Luxembourg	0.75	0.86
Latvia	0.92	0.89
Malta	0.84	1.03
Netherlands	0.45	0.65
Norway	0.20	0.31
Poland	1.02	1.05
Portugal	0.76	1.72
Romania	0.42	0.54
Sweden	0.73	1.06
Slovenia	1.54	2.46
Slovakia	0.81	1.17
United Kingdom	0.43	0.40

Source: Statistics Austria

Table 9

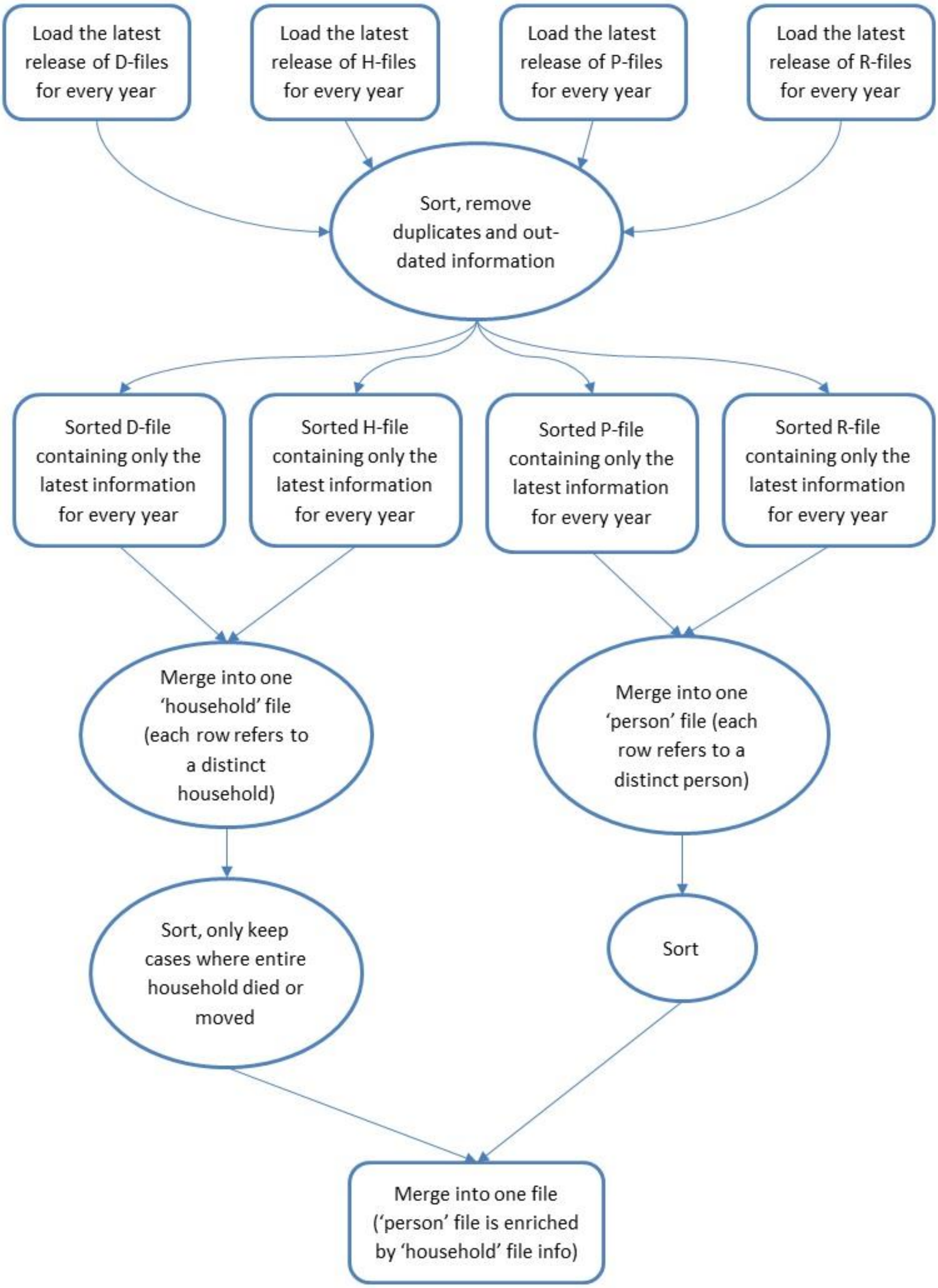
Age-specific death counts and relative frequencies of deaths

Age at First Survey		Number of Observations	Number of Deaths	Percent Deaths
All				
	16 – 19	77,200	233	0.30 %
	20 – 24	80,676	234	0.29 %
	25 – 29	75,796	244	0.32 %
	30 – 34	81,722	228	0.28 %
	35 – 39	88,573	279	0.31 %
	40 – 44	95,739	432	0.45 %
	45 – 49	97,498	618	0.63 %
	50 – 54	95,797	1,013	1.06 %
	55 – 59	91,433	1,260	1.38 %
	60 – 64	82,000	1,613	1.97 %
	65 – 69	67,917	2,013	2.96 %
	70 – 74	57,743	2,714	4.70 %
	75 – 79	47,517	3,830	8.06 %
	Sum	1,039,611	14,711	
Female				
	16 – 19	37,466	106	0.28 %
	20 – 24	39,214	94	0.24 %
	25 – 29	37,990	105	0.28 %
	30 – 34	41,711	81	0.19 %
	35 – 39	45,982	123	0.27 %
	40 – 44	49,994	187	0.37 %
	45 – 49	50,563	246	0.49 %
	50 – 54	49,679	382	0.77 %
	55 – 59	47,590	483	1.01 %
	60 – 64	43,146	626	1.45 %
	65 – 69	36,456	807	2.21 %
	70 – 74	32,039	1,138	3.55 %
	75 – 79	27,600	1,949	7.06 %
	Sum	539,430	6,327	
Male				
	16 – 19	39,733	127	0.32 %
	20 – 24	41,460	140	0.34 %
	25 – 29	37,803	139	0.37 %
	30 – 34	40,008	147	0.37 %
	35 – 39	42,586	156	0.37 %
	40 – 44	45,742	245	0.54 %
	45 – 49	46,934	372	0.79 %
	50 – 54	46,113	631	1.37 %
	55 – 59	43,842	777	1.77 %
	60 – 64	38,853	987	2.54 %
	65 – 69	31,460	1,206	3.83 %
	70 – 74	25,703	1,576	6.13 %
	75 – 79	19,917	1,881	9.44 %
	Sum	500,154	8,384	

Source: Statistics Austria

Figure 1

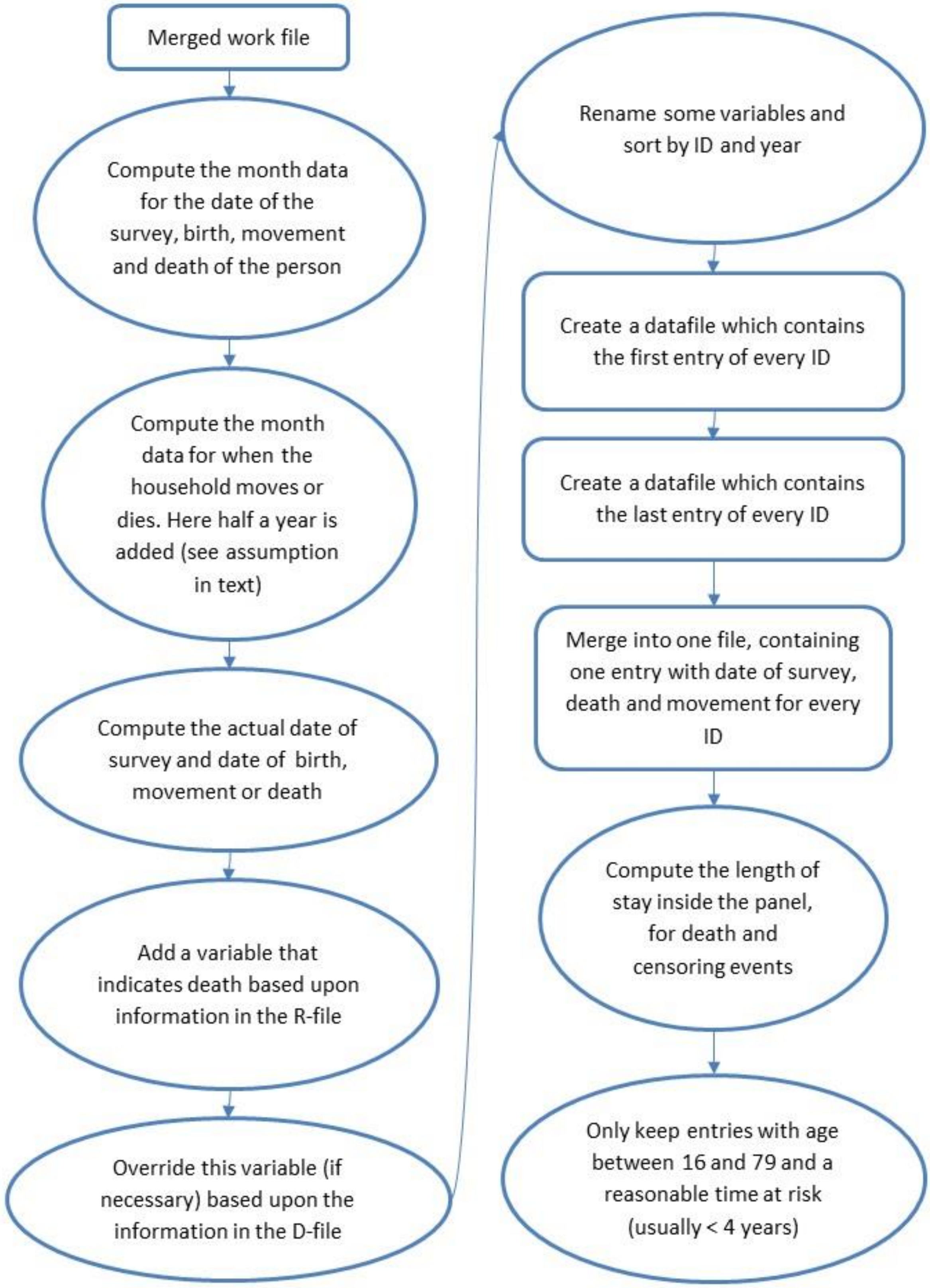
Data query



Source: Statistics Austria

Figure 2

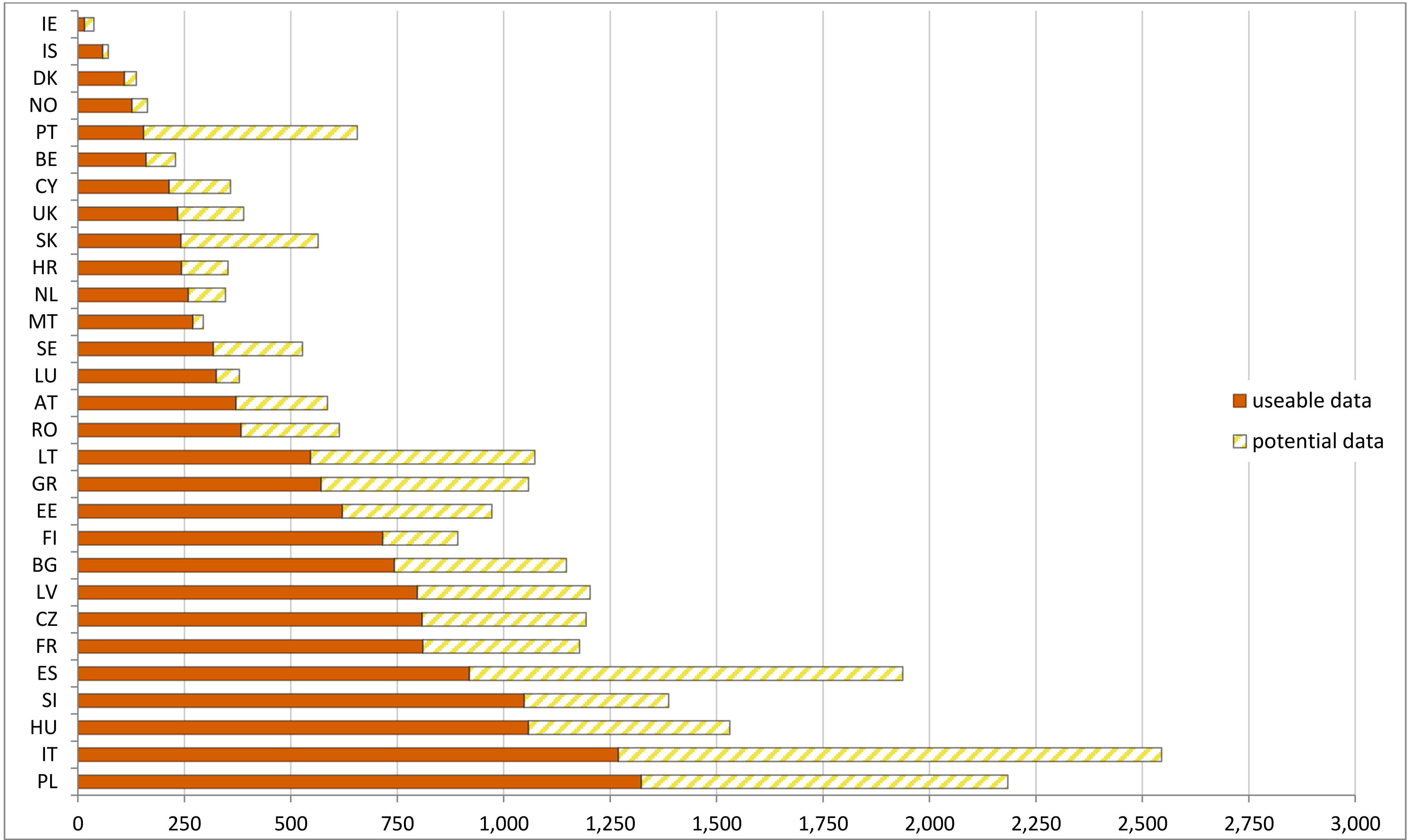
Data editing



Source: Statistics Austria

Figure 3

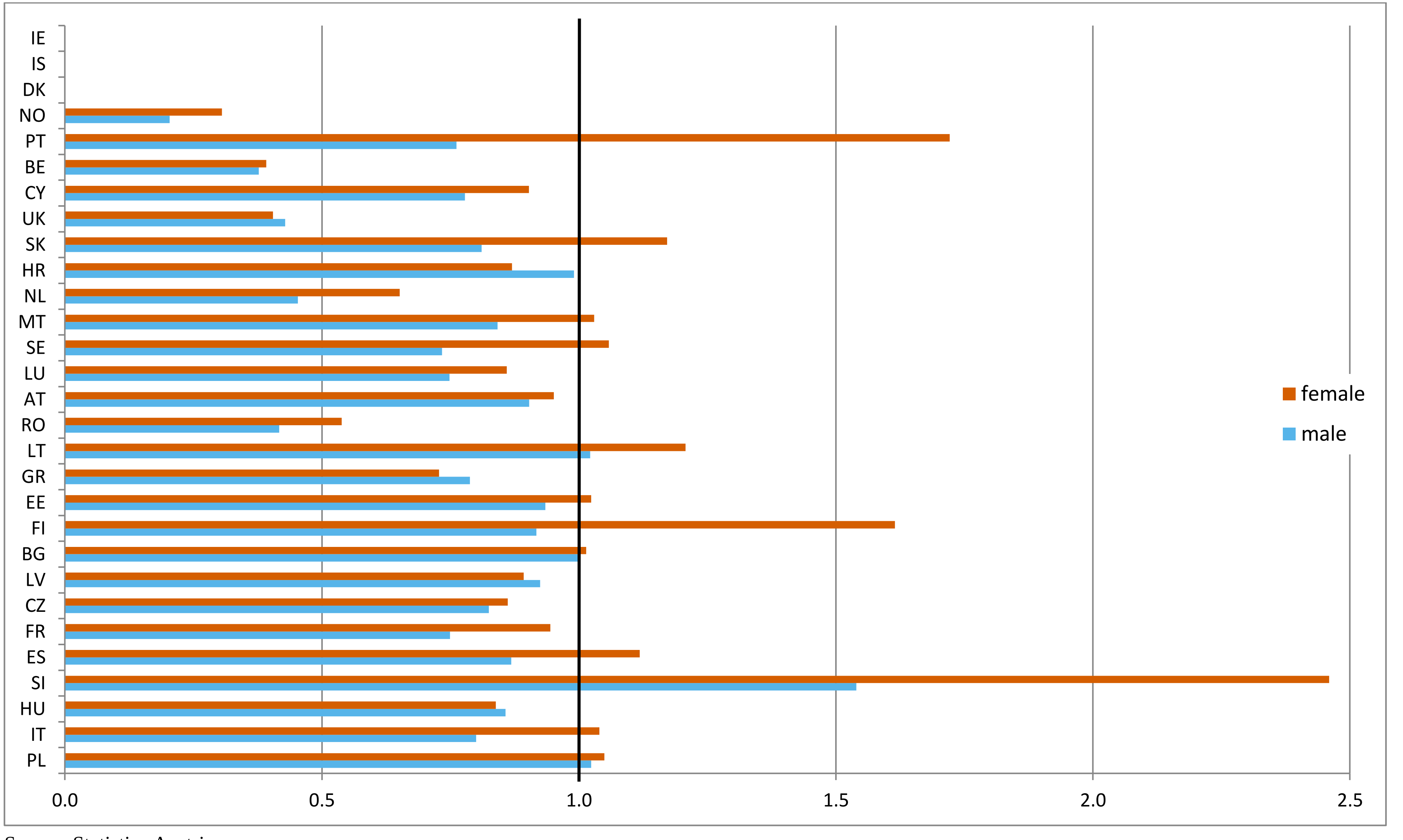
Useable and potential death count per country



Source: Statistics Austria

Figure 4

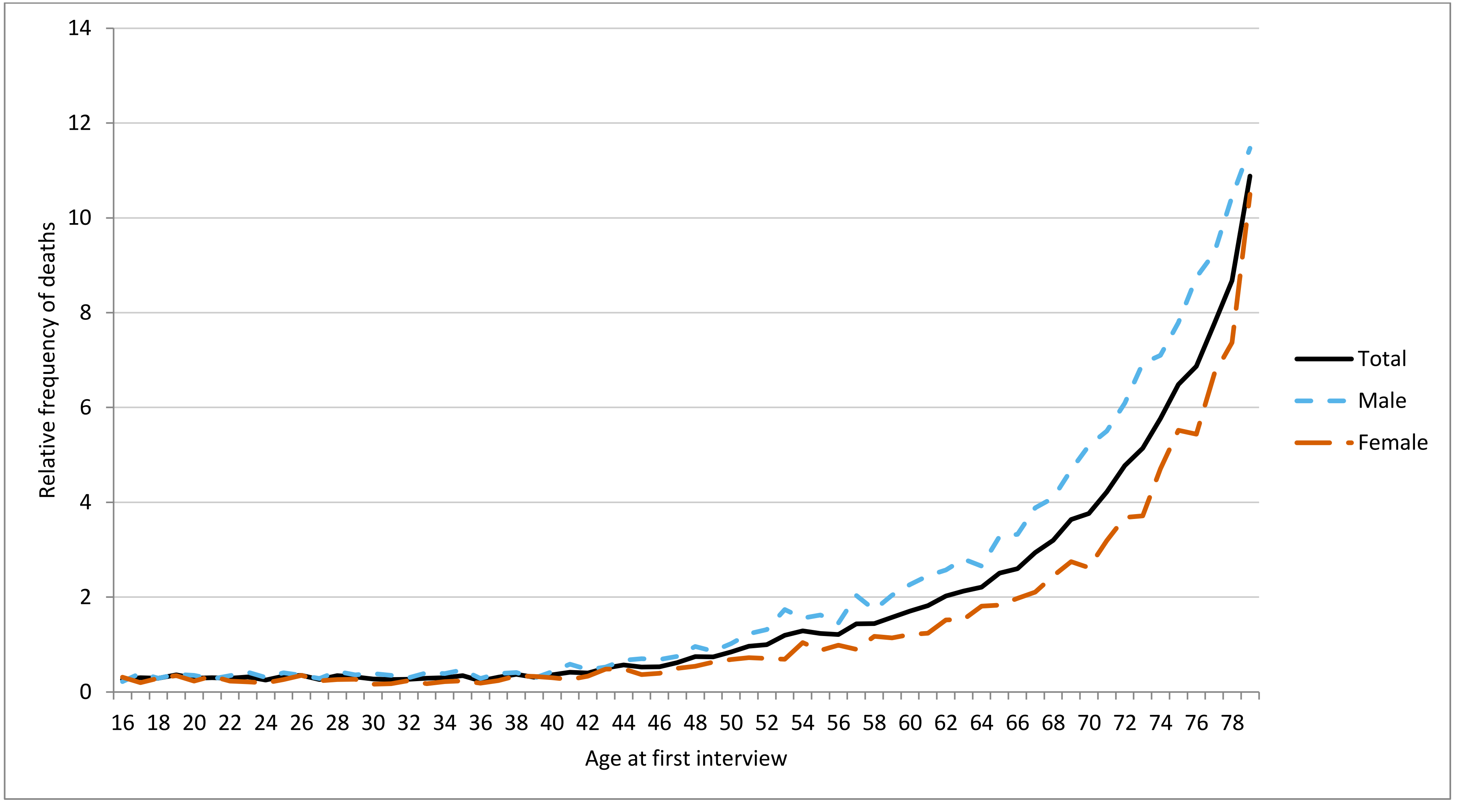
Cumulative relative mortality ratio after 3 years of follow-up



Source: Statistics Austria

Figure 5

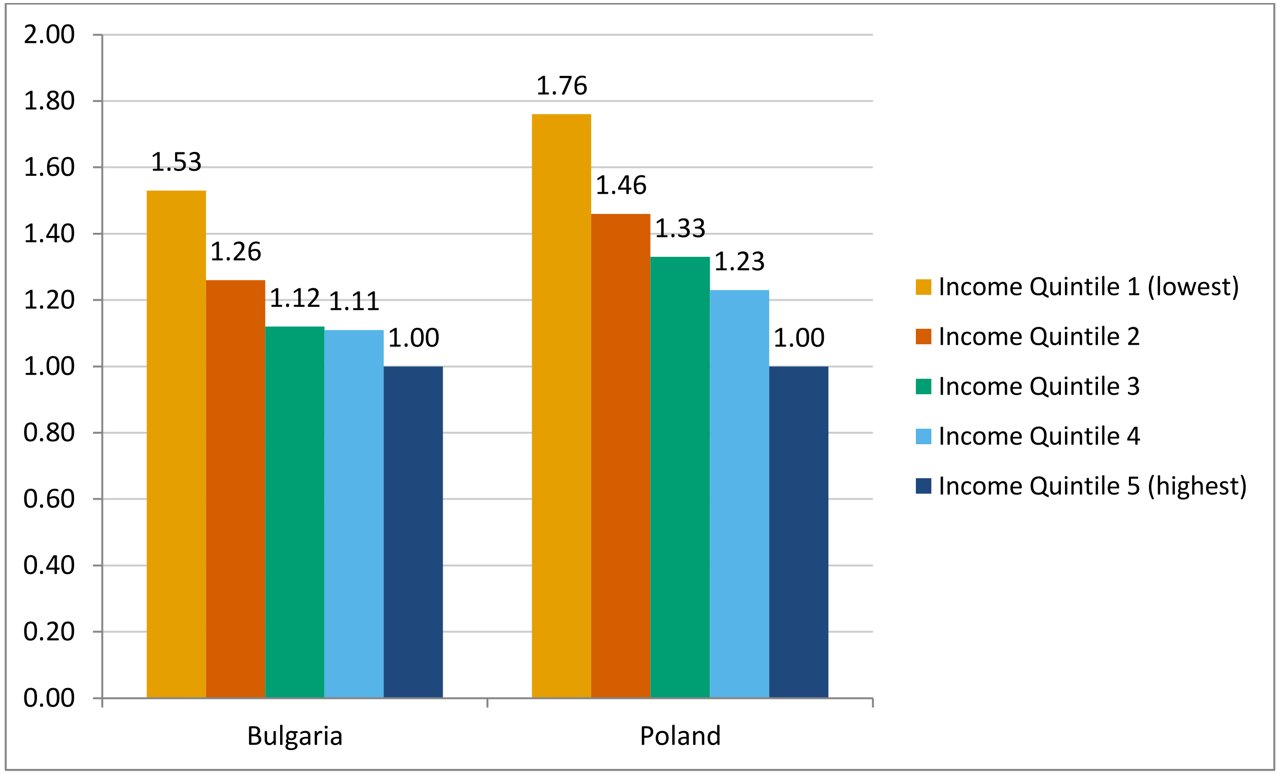
Relative frequencies of deaths



Source: Statistics Austria

Figure 6

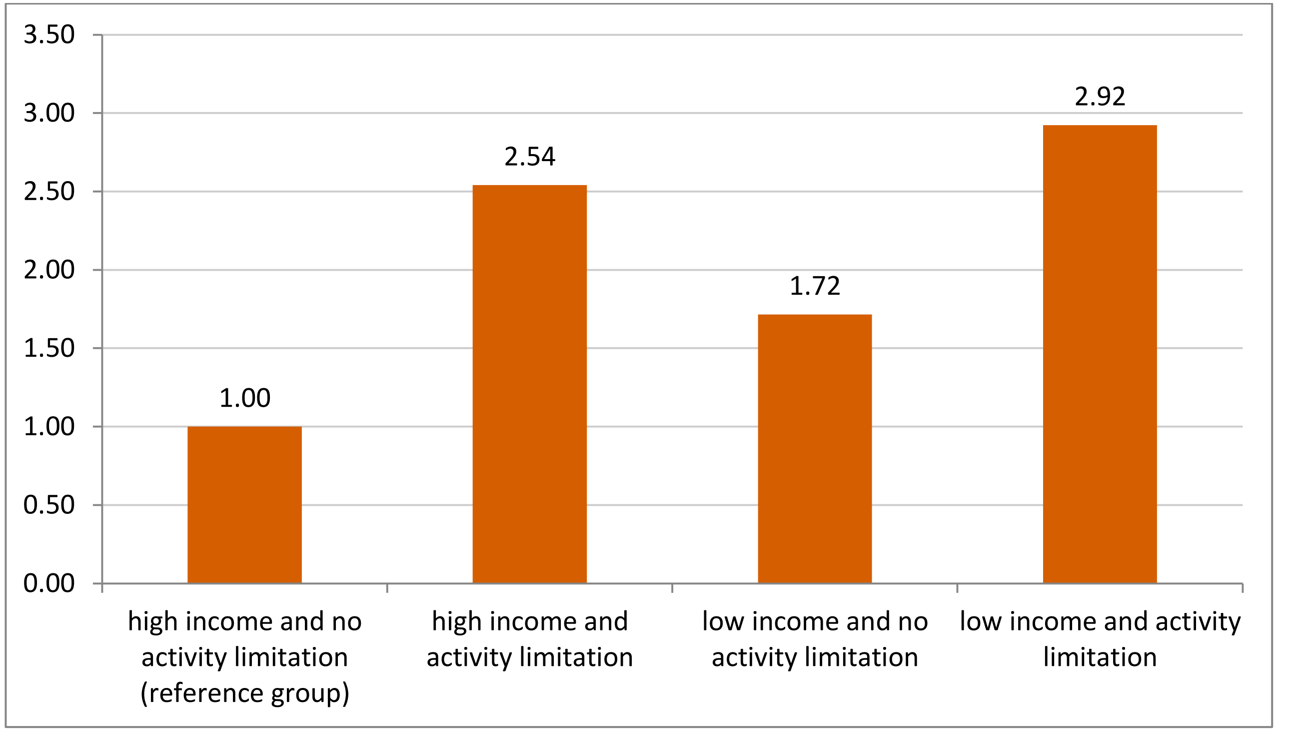
Mortality hazard ratios by income quintile



Source: Statistics Austria

Figure 7

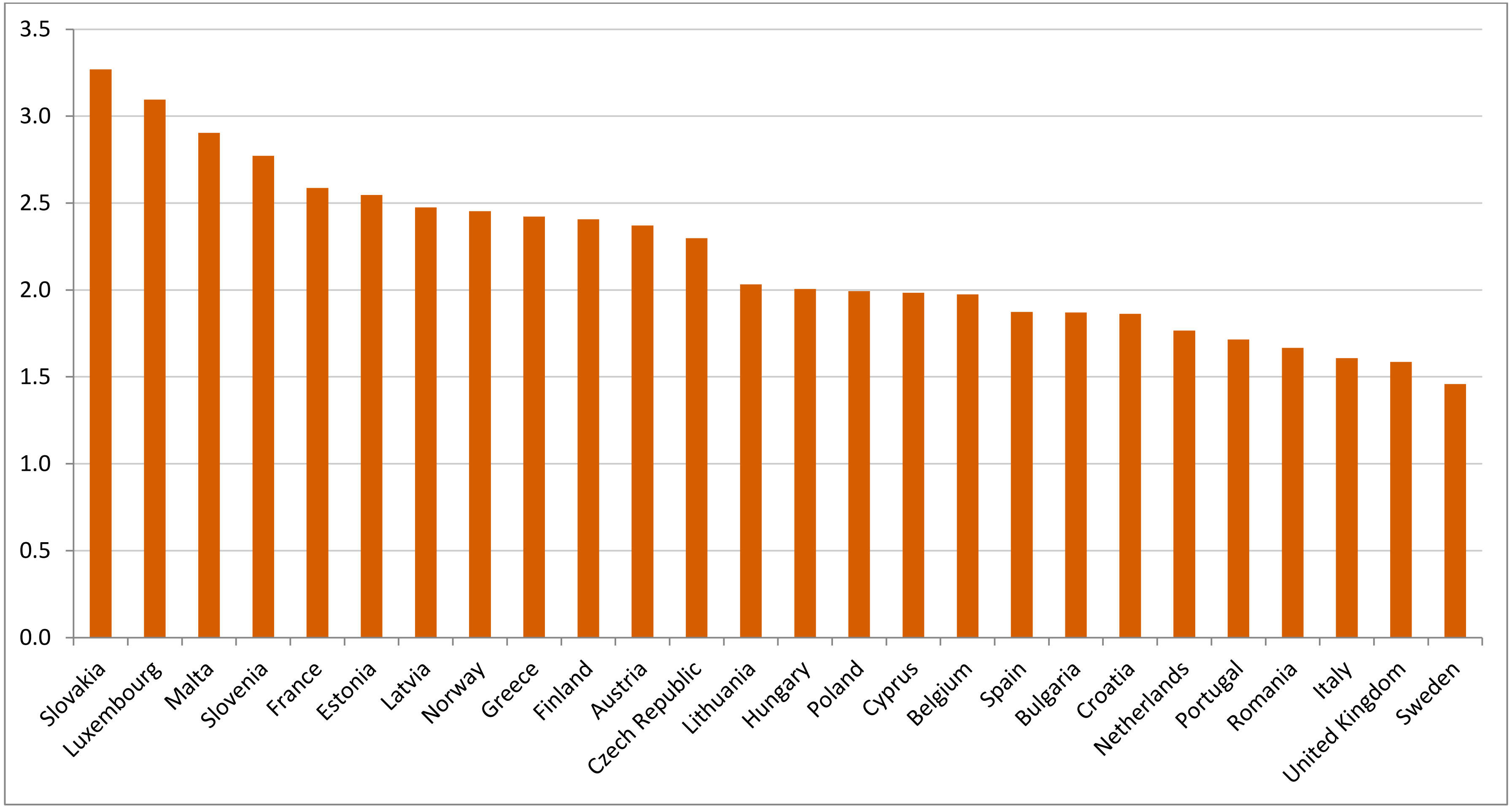
Mortality hazard ratios for Polish males



Source: Statistics Austria

Figure 8

Mortality hazard ratios of low educated males (ISCED 0-2 compared to ISCED 5-6)



Source: Statistics Austria

STATISTICS AUSTRIA
The Information Manager

We understand that the longitudinal component of EU-SILC records when a survey respondent died since the previous wave (target variables DB110, RB110, RB140, RB150).

Before we do comparative analysis on mortality among the poor, please help us to understand the quality of this information by answering the follwing questions:

Your responses refer to which country? ______________________________

(1): How is the information on a respondent's death obtained? (Multiple answers possible)

- ☐ from the interviewer
- ☐ from other household members
- ☐ from a written source (response to a letter /notification)
- ☐ linkage with external data sources, namely ______________________________
- ☐ other, namely ______________________________

(2): All in all, how would you rate the validity of the information about respondents' deaths in the target variables transmitted to Eurostat?

- ☐ very good
- ☐ good
- ☐ adequate
- ☐ poor
- ☐ very poor

(2.1): What are the reasons for your rating?

(3): In case of death of a survey respondent, do you record any additional information which is not included in the target variables transmitted to Eurostat?

☐ Yes. ◯ No.

(3.1): If "Yes.", which additional information is recorded?

(4): In your country, is it theoretically possible to link EU-SILC microdata with mortality information from national death registers (for example, via a unique personal ID variable available in both datasets)?

- ☐ Yes, this has already been done.
- ☐ Yes, but so far it has not been done.
- ☐ Technically possible, but not allowed for legal reasons.
- ☐ No, not possible.

(5): Do you have any further comments or ideas that may help us in regard of mortality analyses from EU-SILC longitudinal data?

(6): Who should we contact in case of further inquiries?

Thank you very much for your cooperation!

You can easily submit your answers using the red button "Send by E-Mail" or write to tobias.goellner@statistik.gv.at

If you have any questions please do not hesitate to contact us.

Send by E-Mail