European Commission

International Agency for Research on Cancer
World Health Organization

Erasmus MC

EU-CanIneq Methodology

to estimate education inequalities in cancer mortality in 2015–2019 for 27 EU Member States, Iceland and Norway

1. Introduction

This document provides an overview of the data and approaches used within the EU-CanIneq project to estimate educational inequalities in cancer mortality (overall and for six specific cancer types) in the period 2015-2019 for 27 European Union member states (MS) plus Iceland and Norway. The project is based on the Erasmus MC Inequality in Health (ERAINHE) dataset (1), which provides harmonized and comparable data across multiple European countries and time periods. To address geographical and temporal gaps, extrapolation methods were developed and applied, extending the estimates beyond the core dataset. This process also integrates information from publicly available databases, including GCO/IARC, Eurostat, and the WHO Mortality Database.

2. Data sources

2.1.The Erasmus MC Inequality in Heath (ERAINHE) dataset

The dataset. The ERAINHE dataset consists of cancer mortality data by education level and sex covering different time periods for individuals aged 40-79 years across 16 European countries, as shown in Table 1. The ERAINHE dataset is built upon a network of international collaborators from various research and statistical institutes across Europe, coordinated by the Erasmus MC. Most data stem from a longitudinal mortality follow-up at the individual level, after a census or individual linkage of national registries and provide information on all-cause mortality and specific causes of death by level of education. For four countries, unlinked data sources were used for deaths and population-at-risk. For all countries, whole-population data are available, except for France where a representative 1% sample was used. The Erasmus MC received the data from national providers and harmonized them. Data is updated as new information is collected and integrated into the dataset, following quality checks and harmonization. The dataset contains death counts and person years both by age group, sex and education level. Age-standardized mortality rates (ASMR per 100,000 individuals) are calculated using the European Standard Population (2).

Depending on the availability of the data, the 16 countries with available data in the ERAINHE dataset could be classified into three main groups, as also reported in Table 1:

A. Countries with at least 3 recorded observations over different periods of time.
B. Countries with 1 or 2 recorded observations only.
C. Countries with no observations for certain cancer sites but with information on total cancer, i.e., France.

Table 1. Overview of observed data on cancer mortality by education level, ages 40–79 years (more details in the Appendix)

	Periods available	Group
Finland	1990-1995; 1995-2000; 2000-2005; 2005-2010; 2011-2015; 2016-2018	A
Sweden	1991 -1994; 1995-1999; 2000-2004; 2005-2009; 2010-2014; 2015- 2017	A
Denmark	1995-1999; 2000-2004; 2005-2009; 2010-2014; 2015-2019	A
Belgium	1991-1997; 2001-2006; 2006-2011; 2011-2015; 2016-2017	A
France	1990-1995; 1995-1999; 1999-2004; 2004-2007	C*
Austria	1991-1992; 2001-2002; 2011-2013; 2013-2014; 2015-2019	A
Italy	2012-2015; 2019; 2020	A
Spain	2002-2006; 2007-2011; 2012-2015; 2016-2019	A
Estonia	2000-2005; 2006-2011; 2012-2015; 2016-2019	A
Lithuania	2001-2005; 2006-2009; 2011-2014; 2015-2019	A
Poland	2001-2003; 2010-2012	B
Czechia	1998-2003	B
Hungary	1988-1991; 1999-2002; 2010-2012; 2015-2019	A
Slovenia	1991-1995; 2002-2006	B
Norway	1990-1995; 1995-2001; 2001-2006; 2006-2009	A
Slovakia	2013-2014; 2015-2019	B

*Given variation in the availability of observed data for the different cancer sites, France was considered as group C. Data is observed for total cancer and lung cancer. For breast cancer data were available but with substantial random fluctuations, and therefore a decision was taken not to use them, while for prostate, cervix, colorectum and stomach cancers data were not available.

Data classifications. Data were available for all cancer types together, as well as for breast, cervix, colorectum, lung, prostate and stomach cancer, as classified in the International Classification of Disease (ICD) 10th revision (1). Educational level was used as an indicator of socioeconomic position and measured at the individual level. To ensure maximum comparability of data between countries, education levels were categorized according to the International Standard Classification of Education (ISCED) [1] as "Primary"(ISCED 0–2: Early childhood, primary, and lower secondary education), "Secondary" (ISCED 3–4: Upper secondary and post-secondary non-tertiary education), and "Tertiary" (ISCED 5–8: Short-cycle tertiary, Bachelor's, Master's, and Doctoral or equivalent levels) (3). ASMR per 100,000 person-years were calculated using the European Standard Population year 2013 (2) to account for the different age structures of the populations across countries and educational groups.

2.2. Other sources of data used

2.2.1. Global Cancer Observatory (GCO) – International Agency for Research on Cancer (IARC)

The GCO-IARC (4) is an interactive web-based platform presenting global cancer statistics to inform cancer control and cancer research. From GCO, it is possible to obtain historical, current and future estimates of cancer incidence and mortality over different countries and geographical regions for different cancer types. The GCO database was used to calculate the rate of change in cancer mortality rates for countries and cancer types where this information was not available in the ERAINHE database (methods B and C; see Section 3.1).

2.2.2. World Health Organization (WHO) Mortality Database

We retrieved data on cancer death count by country, age and sex from the WHO mortality database (5), which provides death registration data for over 120 countries, obtained from countries' annual data by age, sex, and complete ICD code. The raw data files for each year in the period 2015-2019 were collected and the yearly death counts by sex, age group over the age range 40-79 years, and for the relevant cancer sites were obtained for all countries with no data in the ERAINHE database; death counts were used to compute ASMR using the European Standard Population. These were also obtained for countries with observed data in the ERAINHE dataset and used to calibrate the final estimates (see Section 3.3).

2.2.3. Eurostat database

We retrieved data on population by country, age and sex from the Eurostat database (6) which provides population data at country level, by sex and age in 5-year age groups. Country, age and sex-specific population data were obtained for each year in the period 2015-2019 for the age range 40-79 years and for all countries.
Population by education attainment level were also retrieved from the Eurostat database (7) which provides the percentage distribution of the population by educational attainment level, sex, and over different age groups. Data are available from 2013 to 2022 and according to 6 different levels of the ISCED, which were grouped in three categories: ISCED 0-2, ISCED 3-4 and ISCED 5-8. As data for the age group 40-79 years are not available, the distribution of population by education level in the age group 45-79 years was used, and assumed as a proxy. Population estimates from the Eurostat database were used together with the WHO data both to compute mortality rates to apply the "back-calculation" method in countries with no data in the ERAINHE dataset (see Sections 3.2), and to calibrate estimates for countries with observed data in ERAINHE (see Section 3.3).

[1] ISCED: 0 (early childhood), 1 (primary), 2 (lower secondary), 3 (upper secondary), 4 (post-secondary non-tertiary), 5 (short-cycle tertiary), 6 (bachelor's or equivalent), 7 (master's or equivalent), and 8 (doctoral or equivalent)

3. Methods

3.1.Estimation of cancer mortality by educational level for countries with available data in the ERAINHE dataset

The choice of the method for different countries depended on the number of available observations in the ERAINHE dataset and followed the classification into the three groups defined above (section 2.1).

Method for Group A:
ASMRs were computed and used for 2015-2019 if data was available for this period. When only data for earlier periods (for at least three periods) were available, regression models were applied to the ASMRs over calendar time, projecting rates beyond the last observed period and assuming the continuation of past trends until 2015-2019. Separate models were used for each country, sex and education level for total cancer and for the six different cancer types. A log-linear model was chosen when trends were decreasing, to prevent the projections from yielding negative rates. Conversely, in case of increasing trends, a linear model was employed to avoid the exponential increase arising when using a log-linear model. A linear model was also employed when rates were stable (i.e., not significantly increasing or decreasing).

Method for Group B:
For countries with 1 or 2 observations available over time in the ERAINHE dataset, estimates were obtained by calculating the sex-specific average rate of change in mortality overall and for each cancer type using data from GCO-IARC, starting from the last observation period available in the ERAINHE dataset. The sex-specific rate of change estimated for the whole population (not by educational level) was then applied to each educational and sex group assuming the same pace of change over the three educational levels.

Method for Group C:
France was the only country in this group as it had no observations for certain cancer types but available data on total cancer mortality. Mortality estimates for total cancer and for cancer sites with available data were obtained as follows. For total cancer and for cancer sites with available data, we used the method described for group A. For cancer sites with no available data, the share of total cancer mortality attributed to the relevant cancer type was estimated for each educational group and sex, based on the (simple) average proportion obtained from group A countries within the same geographical region in the ERAINHE dataset, in green in Table 2. The estimated share was then applied to the total cancer mortality in the last period with available data, by sex and educational level. Thereafter, the same approach described for Group B was applied.

Table 2. Countries by geographical region

Geographical area	Country
North Europe	Norway, Denmark, Ireland, Sweden, Finland, Iceland
West/South Europe	France, Germany, Austria, Belgium, Netherlands, Luxembourg, Cyprus, Portugal, Spain, Greece, Italy, Malta
Baltic/Central/East Europe	Czechia, Hungary, Bulgaria, Croatia, Latvia, Lithuania, Estonia, Poland, Slovenia, Romania, Slovakia

Group A countries Group B countries Group C countries Countries with no available data

3.2. Estimation of cancer mortality for countries without available data in the ERAINHE dataset

For countries with no observed data in the ERAINHE dataset [2], cancer mortality estimates were obtained using a "back-calculation' (BACK) method. The BACK method consists in borrowing information from countries in the same geographical area, using the ERAINHE dataset, the WHO Mortality Database, and the Eurostat database for the population by age, sex and educational attainment.

The "back calculation" approach consists in the following steps:

1. estimating the geographical area-specific average rate ratio (RR) of cancer mortality for both primary, and secondary as compared to tertiary education (overall and for the six specific cancer types) in 2015-2019 for the three geographical areas (i.e., North, West/South, Baltic/Central/East Europe, table 2), based on group A countries with observed data in the ERAINHE dataset;

2. computing the national average ASMR for cancer (overall and for six specific cancer types) in the age group 40-79 years by sex and country for the period 2015-2019, using death count from the WHO Mortality Database, and both population data from the same period and age range from the Eurostat database;

3. the 2015-2019 ASMRs for each educational group, sex and cancer types, were the calculated applying the estimated area-specific educational relative risks (RRs) to the national average ASMR, while considering the distribution of educational groups in the population, by sex and cancer type.

3.3. Calibration to the WHO database

For countries with observed data in ERAINHE, estimates were calibrated to match the average rates in the period 2015-2019 from the WHO Mortality Database re-calculated to be age-adjusted using the European Standard Population (instead of the World Standard Population). To do so, the mortality rates age-standardised to the European population were obtained by using the number of deaths from the WHO Mortality Database (by country, age and sex, for total cancer and for each cancer type), the population from Eurostat (by country, age and sex) and the European Standard Population. Calibration was then obtained estimating rates for educational groups from the ASMRs considering country-specific RRs (based on the ERAINHE dataset) and the distribution of educational groups in the population by sex. Calibration allowed to ensure internal and external consistency of estimates.

3.4. Sensitivity analysis for the "back-calculation" (BACK) method

To assess robustness of the estimates obtained using the back-calculation (BACK) method, a sensitivity analysis was performed. In this exercise, we compared the estimates for Group A countries (i.e., those based on observed data or short-term projections) with the estimates that would have been obtained by applying the BACK method to the same countries. A leave-one-out sensitivity analysis was performed by sequentially excluding each country from the dataset and applying the BACK method as if data were unavailable for that country. This involved estimating the geographical area-specific average RR of cancer mortality each time excluding a country, and applying the geographical area-specific RRs to the national average ASMR, taking into account the distribution

[2] Germany, Netherlands, Luxembourg, Romania, Portugal, Bulgaria, Latvia, Croatia, Greece, Cyprus, Malta, Iceland, Ireland

of population by education, by sex and cancer type to obtain cause- and sex-specific ASMRs for each educational level. The comparison of these recalculated estimates with the originally estimated values allowed us to assess the robustness of the BACK method and the influence of each country on the overall results. Results from the sensitivity analysis showed that, for the majority of estimates (77%), the relative difference between the values of observed data or obtained through short-term projections and those derived using the BACK method was below 10%. Discrepancies were below 20% for 93% of the estimates and below 40% in almost all cases (99%). A difference greater than 40% was observed for only two estimates, related to cervical and stomach cancer, respectively These cancer types are characterized by relatively low age-standardized mortality rates (ASMRs). Consequently, despite the high relative difference, absolute discrepancies were modest.

3.5. Potential limitations and caution of interpretations

While the EU-CanIneq project is based on most reliable, high-quality data available, the quality may be less than optimal for countries lacking longitudinal data or where recent data were not available. In these cases, statistical models were used to produce the estimates, which may introduce uncertainty. To enhance reliability, these models incorporated information from other countries in the same geographical area (when data were missing for a specific country) “(table2) and leveraged trends in available datasets. ”

Additionally, the time periods analyzed varied by more than a decade across countries with observed data, which may have resulted in an overestimation of the recent RRs and less favorable comparison for those with older data.

Therefore, caution is necessary when interpreting the results, particularly for countries with older and for those with missing data for which the BACK method was used.

Appendix. Sources of data*

Country	Source of the data	Type of data	Notes
Finland**	Preliminary descriptive analyses from joint research project on social determinants of health with Helsinki Institute for Demography and Population Health (TK-53–1490–18)	Longitudinal, individual linkage of national registries	
Sweden**	Preliminary descriptive analyses carried out as part a research project on time trends in social inequalities in health at Stockholm University. The use of Swedish data for this research was provided by the Central Ethical Review Board of Sweden (Dnr Ö 25-2017).	Longitudinal, individual linkage of national registries	
Denmark	Tailored datasets from University of Kopenhagen. Original data from Statistics Denmark	Longitudinal, individual linkage of national registries	
Belgium	Tailoreddatasets from Interface Demography (Vrije Universiteit Brussel), original data from Statistics Belgium	Longitudinal: individual census follow-up in national registers	
France	Tailored dataset form INSERM, original data from INSEE and INSERM	Longitudinal: a 1% representative sample of the population linked with death register	Not including persons born outside France
Austria	Tailored data sets from Statistics Austria	Longitudinal individual linkage census with mortality register.	
Italy	Tailored datasets from Italian National Institute of Statistics (Istat) within the following National Statistical Programme projects: IST-02801 Socio-economic differences in mortality; IST-02844 Feasibility study for a new system to monitor and investigate mortality inequalities.	Longitudinal, individual census population (2011) follow up with National Mortality Register (2012-2015) and the Mortality Inequality Database 2019–2020, based on the integration of National Base Register of Individuals and National Mortality Register (2019, 2020)	

Spain	Tailored datasets from Universidad Complutense de Madrid, original data from National Statistics Institute	Longitudinal: individual census follow-up linked with mortality register	
Estonia	Tailored datasets from National Institute for Health Development, original data from Statistics Estonia	Longitudinal: individual census follow-up linked with mortality register	
Lithuania	Tailored datasets from Lithuanian University of Health Sciences, original data from Statistics Lithuania	Longitudinal, individual census follow-up linked with mortality register	
Poland	Tailored datasets from National Institute of Public Health, National Research Institute, original data from Statistics Poland	Cross-sectional, unlinked: census data and national mortality registry data	
Czechia	Tailored datasets from Charles University, original data from Czech Statistical Office	Cross-sectional, unlinked: census data and mortality registry data	
Hungary	Tailored datasets from Demographic Research Institute at the Central Hungarian Statistical Office	Cross-sectional, unlinked: census data and mortality registry data	
Slovenia	Tailored datasets from the Faculty of Medicine of the University of Ljubljana, original data from the Statistical Office of the Republic of Slovenia	Longitudinal individual census follow-up linked with mortality register	
Norway	Tailored datasets from Norwegian Institute of Public Health, original data from Norwegian Institute of Public Health and Statistics Norway.	Longitudinal, individual linkage of registries	
Slovakia	Tailored datasets from Statistics Slovakia	Cross-sectional, unlinked census and mortality registry data	

* Data were collected and harmonized in close collaboration with national partners in different projects. Projects contributing the data collection and harmonization varied between countries and periods and include: the EU-CanIneq project, funded by the European Commission number 101115691, the CHAIN project funded by the Nordic Research Council (grant number 288638) the LIFEPATH project, funded by the European Commission (Horizon 2020 grant

number 633666); the project “Longer life, longer in good health, working longer? Implications of educational differences for the pension system’, funded by Network for Studies on Pensions, Aging and Retirement (NETSPAR); the SOCINEQ project, funded by the Institut National du Cancer (grant 2018-116); the DEMETRIQ project, funded by the European Commission (grant numbers FP7-CP-FP and 278511); the Euro-GBD-SE (funder: European Commission, grant number 20081309). It is complemented with descriptive analyses carried out as part of joint research projects with Erasmus MC where data could not be included in the ERAINHE dataset.
** Due to data sharing restrictions, Finland and Sweden are not part of the ERAINHE dataset, descriptive analyses using age-standardized mortality rates (ASMR), derived from collaboration with data providers, are included in the current factsheets.

Contact information

IARC: Cancer Inequalities Team, Cancer Surveillance Branch, International Agency for Research on Cancer. eu-canineq.iarc.who.int
European Cancer Inequalities Registry (ECIR): cancer-inequalities.jrc.ec.europa.eu ec-ecir@ec.europa.eu sante-rtd-cancer@ec.europa.eu

References:

1. Vaccarella S, Georges D, Bray F, Ginsburg O, Charvat H, Martikainen P, et al. Socioeconomic inequalities in cancer mortality between and within countries in Europe: a population-based study. Lancet Reg Health Eur. 2023;25:100551.
2. Pace M, Lanzieri G, Glickman M, Zupanič T. Revision of the European Standard Population: report of Eurostat's task force: Publications Office of the European Union; 2013.
3. Statistics Ulf. International standard classification of education: ISCED 2011. Comparative Social Research. 2012;30.
4. Filho AM, Laversanne M, Ferlay J, Colombet M, Piñeros M, Znaor A, et al. The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. International Journal of Cancer. 2025;156(7):1336-46.
5. World Health Organization Mortality Database. Available at https://www.who.int/data/data-collection-tools/who-mortality-database .
6. Eurostat. https://ec.europa.eu/eurostat/databrowser/view/demo_pjangroup__custom_11728949/default/table?lang=en .
7. Eurostat, by education. https://ec.europa.eu/eurostat/databrowser/view/edat_lfs_9903__custom_10784414/default/table?lang=en .