Supplementary Information –

Additional information regarding the mortality data by educational level

Journal: Journal of Epidemiology and Community Health

Jesús-Daniel Zazueta-Borboa[1*], Pekka Martikainen[2,3], José Manuel Aburto[4,5,6], Giuseppe Costa[7], Riina Peltonen[2], Nicolás Zengarini[7], Alison Sizer[8] Anton E. Kunst[9], Fanny Janssen[1,10]

[1] Netherlands Interdisciplinary Demographic Institute – KNAW/University of Groningen, Lange Houtstraat 19, 2511 CV The Hague, The Netherlands.

[2] Population Research Unit, Faculty of Social Sciences, University of Helsinki, Finland

[3] Max Planck Institute for Demographic Research, Rostock, Germany

[4] Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

[5] Leverhulme Centre for Demographic Science, University of Oxford, Oxford, United Kingdom

[6] Interdisciplinary Centre on Population Dynamics, Southern Denmark University, Odense, Denmark

[7] Epidemiology Unit, ASL TO3, Piedmont Region, Grugliascso (TO), Italy.

[8] Department of Information Studies, University College London, United Kingdom.

[9] Department of Public and Occupational Health, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

[10] Population Research Centre, Faculty of Spatial Sciences, University of Groningen, The Netherlands

* corresponding author

Zazueta-Borboa JD, *et al. J Epidemiol Community Health* 2023; 77:421–429. doi: 10.1136/jech-2023-220385

The text below is largely copied from the working document "General description of the mortality data by educational level as part of the Future Longevity Inequalities project". See https://www.futurelongevitybyeducation.com/background-information/ (Password = VICI_info).

We used all-cause mortality data by highest educational level attained (low, middle, high), sex, single age (30+) and single calendar years for England & Wales (1972-2017), Finland (1971-2017) and Italy (Turin) (1972-2019). These data stem from longitudinal follow-up designs in which individual mortality data are linked to individual information regarding educational attainment from population censuses (Finland and Italy) or census subsamples (England & Wales) up to 10 years earlier.

For Finland (1971-2017), the data stem from Statistics Finland, and cover all official residents in Finland (OSF 2020a); for England & Wales (1972-2017), the data stem from the Office for National Statistics Longitudinal Study (ONS-LS), and pertain to a 1% representative sample of the population (Shelton et al., 2019); and for Italy (1972-2019), the data stem from the Turin Longitudinal Study, and cover the population residing in Turin (Costa and Demaria, 1988; Cardano et al., 2004).

In measuring highest educational attainment obtained, we distinguished between low (ISCED 0-2), middle (ISCED 3-4), and high (ISCED 5+) levels according to either the International Standard Classification of Education (ISCED) from 1997 (UNESCO, 1997) or the ISCED 2011 (UNESCO, 2012), which are effectively the same.

For summary information on the data sources, the follow-up design we used in each country, and the resulting aggregate data, see Table A1. More detailed information by country is provided below.

Zazueta-Borboa JD, *et al. J Epidemiol Community Health* 2023; 77:421–429. doi: 10.1136/jech-2023-220385

England & Wales (E&W)

For England & Wales, we used information from the Office for National Statistics Longitudinal Study (ONS-LS), which comprises individual information on demographic and socio-economic variables obtained through the censuses (1971, 1981, 1991, 2001, 2011), as well as individually linked information on life events (including births, deaths, and cancer registrations), for an approximately 1% representative sample of the population of England & Wales (Shelton et al. 2019). More specifically, the ONS-LS includes all individuals born on four undisclosed birthdays per year, and the study is maintained through the annual addition of new-borns and immigrants with the same birth dates (Shelton et al. 2019).

For our analysis, we employed a 10-year follow-up of the ONS-LS sample members aged 20 and older at the time of the census in 1971 (25 April), 1981 (5 April), 1991 (21 April), 2001 (29 April), and 2011 (27 March). We followed these individuals until the next census, or until the date that vital status information was last linked to the LS (currently 31 December 2017).

To ensure that we had complete information on educational attainment, we did not count immigrants until the census and the follow-up period after they arrived in England & Wales. For individuals who emigrated but who also returned during a given follow-up period (= were present at the next census), the person-time for all 10 years was counted (or, if they died, until their date of death). Emigrants who had not returned by the next census (= were not present at the next census and had not died) were excluded from the entire follow-up period. We defined emigration in this way (as not being present at the next census and not having died) because in England & Wales, official emigration statistics with emigration dates are considered far from complete (CeLSIUS, 2021a). Between the 2011 census and 31 December 2017, we could only define, and subsequently exclude, emigrants based on the available data on emigration in the ONS-LS (through the "embarkation" variable), since data for the subsequent census (2021) will not be available until late

2023. However, we corrected for the use of these two different emigration definitions in our data-handling process (see Janssen et al. (in progress)).

The information on mortality within the ONS-LS stems from two sources: first, the ONS Death files were scanned for individuals with an LS birthday, and those files that were linked to an LS member were linked to the LS; and, second, notification was sent to the LS Development Team at the ONS of the deaths of people who were flagged in the NHS Central Register (NHSCR) as LS members. The use of both sources helped to ensure that very few deaths of LS members were missed. However, because it was not possible to quantify the exact number of missed deaths, we followed the recommendation to exclude those LS members who were not traced in the NHSCR (= individuals for whom their events, including their potential death, were not picked up) (CeLSIUS 2021b).

The Longitudinal Study Development Team at the ONS used restricted information on the date of birth and the date of death to transform the individual cohort data into aggregate period data. At the individual level, the counting of the person-days per calendar year started either at the beginning of each year or on the subject's 30th birthday, and ended on 31 December of each year or on the subject's date of death. The data were then aggregated, resulting in all-cause and cause-specific mortality data for individuals aged 30+ by educational level, sex, and age from 1 January 1972 to 31 December 2017.

The education information in the ONS-LS is not available in a consistent manner over time, and only the education information from the 1971, 2001, and 2011 censuses aligns reasonably well with the ISCED data. Moreover, the information on educational attainment is not complete because in the 1971 and the 2001 censuses, individuals who were aged 70 and older or aged 75 and older, respectively, were either not asked or were not required to answer the education questions. See Janssen et al. (in progress) for a more detailed description of these issues, and of how we dealt with them. In implementing the data adjustment, we used the all-cause mortality data by five-year age groups (30-34, …, 90-94, 95+). We then smoothed the adjusted all-cause mortality data (death

counts, personyears) into data by single year of age (30, …, 99, 100+) using the Rizzi et al. (2015) smoothing technique, while selecting age 110 as the final age for the open-ended 95+ age group. In doing so, we maintained the totals not only across age groups, but also across educational groups.

Remote access to the micro-data needed for our project was provided by CeLSIUS (Centre for Longitudinal Study Information and User Support). With their help, we were able to run our data management and data aggregation scripts, and were, in turn, able to run our analysis scripts on the aggregate data.

Finland

We applied a follow-up of the Finnish population aged 25 and older on 31 December of the (census) years 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, and 2015 during the subsequent five years; or, for the last follow-up period, until the last date for which data are available (31 December 2017). Up to 1985, individuals were followed up using data from the official (traditional) census; whereas after 1985, individuals were followed up using data from the annual population registration (= virtual census).

In each of these five-year follow-up periods, we excluded individuals who emigrated during the period and did not return by the end of the period. We did so because no exact emigration dates were available before 1990, which made it impossible for us to track emigrants more precisely while maintaining comparability over time. For individuals who emigrated but came back within the same follow-up period, the person-time was counted for each year of the follow-up period (or, if they died, until their date of death). Immigrants were not immediately counted from the moment they entered Finland during a given follow-up period, but were included in the subsequent follow-up period to ensure that information on their educational level was available. At the individual level, the counting of person-days started at the beginning of each year or on the subject's 30th birthday, and ended on 31 December of each year or on the subject's date of death. After

aggregating the data by calendar year and age, we ended up with age-period data for individuals aged 30+ from 1971 up to 2017.

We obtained remote access to the micro-data through Fiona, and were therefore able to run our analysis scripts on the original data. Because of privacy issues, all individuals in the Statistics Finland data were assigned a birth date on the 15th of the month.

The information on education comes from registers of post-compulsory educational certificates, and could be easily translated to the ISCED codes we distinguished. We determined each individual's educational level at the start of each five-year follow-up period. In doing so, "missings" – indicating that these individuals had no post-compulsory educational certificate – were classified as low educated, because it is explicit that these "missings" had basic education only.

Italy (Turin)

For Italy, we made use of the Turin Longitudinal Study (Costa & Demaria, 1988; Cardano et al., 2004), which includes demographic, socio-economic, and health information on all individuals who have been resident in Turin (the second-largest city in Northern Italy) once or more since January 1971. This information was obtained through individual record-linkage procedures. For our research project, we performed 10-year follow-ups of the Turin population aged 20 and older covered in the population censuses in 1971 (24 October), 1981 (25 October), 1991 (20 October), 2001 (21 October), and 2011 (9 October). More specifically, we performed a follow-up from one census until the next census; and, for the 2011 census, we performed a follow-up until the final observation year (31 December 2019).

For individuals who emigrated but returned during a given follow-up period, the person-time for the whole follow-up period was counted (or, if they died, until their date of death). For individuals who emigrated but did not return during a given follow-up period, the person-time until their emigration date was counted. Immigrants were not counted until the following census (= the following follow-up period) to ensure that we had complete information on their educational level.

To enable us to work with full observation years, we defined the study period for Turin as starting on 1 January 1972 and (currently) ending on 31 December 2019. For our analysis, we made use of the exact dates of birth, death, and emigration obtained through the population registers. Similarly, to transform the individual cohort data into aggregate period data, we used the information from the population registers.

The information on educational attainment comes from the different censuses. The Italian qualifications were grouped into the three ISCED categories of low, middle, and high following the ISCED recommendations (UNESCO 2017). Individuals with missing educational attainment were deleted from the follow-up period. The percentage of people with missing educational attainment was less than 1%, except in the last follow-up period, for which the percentage of people with missing educational attainment was around 3%.

Zazueta-Borboa JD, *et al. J Epidemiol Community Health* 2023; 77:421–429. doi: 10.1136/jech-2023-220385

References

Cardano, M., Costa, G., Demaria, M. (2004). Social mobility and health in the Turin longitudinal study. *Social Science & Medicine,* 58(8), 1563-74. doi: 10.1016/S0277-9536(03)00354-X.

CeLSIUS (2021a). *Events Guide*, Centre for Longitudinal Study Information and User Support (CeLSIUS), accessed 09/11/2021, <https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/health-and-social-surveys-research-group/studies-21>

CeLSIUS (2021b). *Mortality Guide*, Centre for Longitudinal Study Information and User Support (CeLSIUS), accessed 09/11/2021, <https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/health-and-social-surveys-research-group/studies-46>

Costa, G. & Demaria, M. (1988). Un sistema longitudinale di sorveglianza della mortalità secondo le caratteristiche socio-economiche, come rilevate ai censimenti di popolazione: Descrizione e documentazione del sistema. *Epidemiologia e Prevenzione*, 36, 37–47.

Janssen, F., Van Hemelrijck, W.M.J., Kagenaar, E., Sizer, A. (submitted). Examining long-term mortality trends by education level for England and Wales: solving data issues.

Official Statistics of Finland (OSF)(2020a). *Documentation of statistics Deaths*. Helsinki: Statistics Finland https://www.stat.fi/en/statistics/documentation/kuol. Accessed 12/02/2021.

Office for National Statistics (2019). ONS Longitudinal Study – England and Wales. Released 11 June 2019, ONS SRS Metadata Catalogue. doi: 10.57906/z9xn-ng05.

Rizzi, S., Gampe, J., Eilers, P.H. (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology,* 182, 138–147. doi: 10.1093/aje/kwv020.

Shelton, N., Marshal, C.E., Stuchbury, R., Grundy, E., Dennett, A., Tomlinson, J., Duke-Williams, O., Xun, W. (2019). Cohort Profile: the Office for National Statistics Longitudinal Study (the LS). *International Journal of Epidemiology*, 48(2): 383–384. doi: 10.1093/ije/dyy243.

UNESCO (1997). *International Standard Classification of Education - ISCED 1997*. Paris: UNESCO Institute for Statistics.

UNESCO (2012). *International Standard Classification of Education - ISCED 2011*. Paris: UNESCO Institute for Statistics.

UNESCO (2017). *International Standard Classification of Education.* Paris: UNESCO Institute for Statistics. https://isced.uis.unesco.org/visualizations/. Accessed 10/12/2020.

Table A1. Summary information on the data sources used and the follow-up designs applied, and the resulting aggregate data, by country

	England &Wales	**Finland**	**Italy (Turin)**
Overall source of data	ONS Longitudinal Study	Statistics Finland	Turin Longitudinal Study
Follow-up population	Sub-sample (1%) of the E&W population aged 20 and older	Finnish population aged 25 and older	Turin population aged 20 and older covered in the population censuses
Length of follow-up	10 years	5 years	10 years
(Underlying) source of educational data	The different censuses (25-04-1971, 05-04-1981, 21-04-1991, 29-04-2001, 27-03-2011)	Registers of post-compulsory education certificates (1 Jan 1971, 1976, …. , 2016)	The different censuses (24-10-1971, 25-10-1981, 20-10-1991, 21-10-2001, 9-10-2011)
(Underlying) source of population at risk	ONS Longitudinal Study	Up to 1985 based on census data (1970, 1975, 1980, 1985). After 1985 based on annual population registration.	Population registry of the National Institute of Statistics
(Underlying) source of all-cause mortality data	ONS Death files and the NHS Central Register	Population Information System of the Population Register	Municipality Population Registry of the City of Turin
Treatment of immigration	Only considered when they were present at the start of the follow-up period		
Treatment of emigration	Individuals who emigrated and did not return during a given follow-up period were excluded from that follow-up period.	Individuals who emigrated and did not return during a given follow-up period were excluded from that follow-up period.	For individuals who emigrated and did not return during a given follow-up period, the person-time until their emigration date was counted.
Treatment of temporal emigration	For individuals who emigrated and came back within the same follow-up period, the person-time was counted for the full follow-up period.		
Source of migration data	Based on determining whether subjects in the different follow-ups – and who did not die – were still present in the subsequent census.	Up to 1985 => based on census data (in line with E&W). From 1986 onwards => based on annual population registration.	Population registry of the National Institute of Statistics
Data based on exact date of birth and death?	Yes.	No. Everyone was assigned a birth date on the 15th of the month	Yes
Additional information	We made necessary adjustments to the data. See Janssen et al. (in progress).	NA	NA
Resulting aggregate population data:			
Single years of age	30-100+	30-100+	30-95+
Yearly data	1972-2017	1971-2017	1972-2019
N of deaths (30+)	257,413 (original); 247,425 (adjusted)	2,191,480	419,637
N of personyears (30+)	14,372,539 (original); 14,358,549 (adjusted)	142,033,351	27,665,912