Notes to 'Annual deaths by cause, age and sex in Scotland, 1855-1949'.

1. Introduction

This dataset is a machine-readable transcript of the counts of deaths by age, sex and cause recorded in the annual Reports of the Registrar-General for Scotland, covering the years 1855-1949. Cause of death data were collected for Scotland from the introduction of civil registration in 1855.[1] Although this post-dated the English system of civil registration, the delays in developing the Scottish system avoided the inconsistencies evident in the early years of the English system, and the Scottish data were published annually in a continuous series from 1855. From 1950 the series is available through the WHO.[2] Together with the WHO database this series constitutes the second longest cause of death series for any national population available in machine-readable form. The longest series is that for England and Wales, from 1848.[3]

The early Scottish registrar-generals adopted a different nosology from that developed in Enlgand and Wales under William Farr, and this makes comparison with the English statistics of particular interest in understanding the impact of nosology on apparent trends in cause-specific mortality.[4] Critically, the database provides deaths by individual cause, allowing users to estimate transfers of deaths between disease groups, and to aggregate the data in multiple ways. The dataset also provides a national standard for Scottish micro-demographic studies. Scottish vital registration and census records provide far greater detail than those for England and Wales and therefore make possible record linkage studies of the impact of individual-level characteristics on life chances, that are simply unviable with English records.[5]

The present dataset has a number of advantages and disadvantages, set out below.

Advantages.
1. The data are available annually, enabling the construction of long-run series of age-specific and cause-specific mortality, and the detection of inter-year variation.
2. The high resolution of the data by cause allows users to create their own aggregated categories of causes, and to test to some extent whether changes over time in mortality by cause are artifactual. The tables that have been made machine-readable in this dataset

---

[1] See Cameron (2007) for an account of the development of death registration in Scotland. See also the Wellcome-funded project 'The Scottish way of birth and death'
http://www.gla.ac.uk/schools/socialpolitical/research/economicsocialhistory/historymedicine/scottishwayof birthanddeath/
[2] WHO mortality database: http://www.who.int/healthinfo/morttables/en/index.html
[3] 1848-1900: ESDS SN5705; 1901-2000: 'The twentieth century mortality files – 20th century deaths' and subsequent years available from the Office of National Statistics
(http://www.ons.gov.uk/ons/rel/subnational-health1/the-20th-century-mortality-files/20th-century-deaths/index.html)
[4] See Crowther (2006)
[5] See http://www.hpss.geog.cam.ac.uk/research/projects/victorianscotlanddemography/,
http://www.hpss.geog.cam.ac.uk/research/projects/doctorsdeaths/).[5]

have been used previously to follow individual causes of death, but the absence of a fully digitized dataset meant that there was no systematic investigation of the reallocation of deaths between causes.

3. The data are available for single years of age for ages 0-4 before 1911, and for 5 year age groups at ages 5+ for all years except the period 1911-1924 (when deaths were reported in ten year age groups for ages 15+). Therefore it is possible, by smoothing annual cause-specific death rates by age, to create long-run cohort data by single year of birth. The spatially aggregated nature of the data is an advantage in this regard, because biases introduced by migration will be less significant at the national level than would be the case for smaller units, however the effects of emigration in particular must be borne in mind. Infant deaths were also reported for three age groups within the first year of life before 1911, providing higher resolution of age- and cause-specific infant mortality than is available for England and Wales for most of this period.

4. The data are available by sex, making it possible to calculate sex ratios for every age and cause. This is of intrinsic interest, and is also a useful diagnostic tool when investigating changes in mortality over time. A change in the sex ratio may indicate a change in the criteria for the allocation of deaths to a particular cause. The sex ratio can also be useful in attempting to reallocate deaths from obscure or 'ill-defined' causes to other categories, where causes have a distinctive sex ratio.

5. Annual estimates of the population by age and sex are available for Scotland, making it possible to calculate annual age-specific mortality rates.[6]

Disadvantages.

1. The data lack spatial resolution, representing only the aggregate experience of disparate populations. This problem can be addressed to some extent by using the dataset in conjunction with other spatially disaggregated datasets of deaths by age and cause published by the Scottish Registrar General.

2. The data lack temporal resolution beyond the annual level, so seasonal patterns in causes of death cannot be observed. For some causes seasonality can be inferred from local data collected on a weekly or quarterly basis, although the latter are usually not available by age or sex.

3. The dataset suffers from all the problems associated with cause of death data (misspecification of cause, changes in quality of data over time, changes in nosology), and with historical data (in particular there is some under-recording of infant deaths[7], and some misrecording of ages). In addition the data are highly aggregated and lack any information about cause-specific mortality by occupation or income.

2. The dataset.

The dataset contains annual cause-specific mortality data by age and sex for Scotland for the years 1855-1949, transcribed from the Annual Reports of the Registrar-General of Scotland. The dataset consists of a single file spanning the period 1855-1949 ('dth_1855-1949.txt'). Deaths are enumerated by age and sex for each cause, and each cause is categorised according to the classification systems used in the Annual Reports.

The data were transcribed from the printed Reports into Excel spreadsheets and then compiled and checked for errors (see section 5) in STATA. The dataset is available as a tab-delimited text file ('dth_1848-1900.txt') that can be imported into Excel, Access or statistical software

---

[6] Available from the Human Mortality Database: http://www.mortality.org/
[7] See for instance Davis (2009) regarding the registration of stillbirths in Scotland.

packages. The dataset structure is simple. Each line contains a single cause of death for a particular year, age group and sex. There are 11 variables, described in section 4.

Age-specific mortality rates can be calculated from the data using the population estimates for Scotland in the Human Mortality database.[8]

3. The classification systems.

There were a number of changes in the nosology employed in the Annual Reports over the period 1855-1949. The Scottish Registrar-Generals devised a system that differed from those used in England and Wales, and the classification schemes are not directly comparable before 1911, although comparison should prove instructive. The nosological scheme employed was revised several times before 1911, in 1877, 1883 and 1901. From 1911 Scotland adopted the International Classification of Deaths (ICD) coding system. The ICD scheme underwent periodic revisions, in 1921, 1931 and 1941.[9]

Users are advised to check for consistency of causes and categories of causes over time, when using the dataset to construct time series. While a cause may appear unchanged in name between coding periods, there were sometimes abrupt shifts in the classification of types of deaths, resulting in a sudden decrease or increase in the number of deaths under a particular heading. Users should also be aware that even a gradual change in death rates over time for a given cause does not necessarily indicate a genuine change in mortality, since medical trends could cause a gradual reallocation of deaths between categories (the most problematic, especially in the case of infants and older adults, being the reallocation of deaths from 'ill-defined' to distinct causes). One advantage of the present dataset is that the availability of cause-specific data at a high level of resolution allows the user to check for evidence of reallocation of deaths between individual causes.

4. The variables.

The file 'dth_1855-1949.txt' contains 11 variables, described below.

cause (string)

Cause of death. The Annual Reports contain a fairly comprehensive list of causes by age and sex for Scotland. The terminology and classification of causes of death was not uniform over the period, and the various classificatory schemes are described in section 3. *When constructing series, users should investigate absences of causes to check whether the absence represents a change in classification (for instance a new variant of name, or deletion of that cause upon periodic reclassification – see section 3) or an absence of deaths in that year (indicated by presence of the cause in subsequent years).* In addition to individual causes, total deaths were reported as 'all causes'.

year (integer)

Calendar year of death.

sex (integer)

Male (1) or female (2).

agegp (string)

---

[8] See footnote 6.

[9] See http://www.scotpho.org.uk/publications/overview-of-key-data-sources/scottish-national-data-schemes/deaths for an introduction to the ICD schema.

Age groups. The age groupings are not consistent over the period. Deaths were recorded for single year age groups for ages 0-4 for every year 1855-1910. In addition deaths under one were broken down into ages 0-2 months, 3-5 and 6-11 months for the years 1855-1910. Deaths were recorded in five year age groups for ages 5+ except in the years 1911-1924, when deaths were reported in ten year age groups for ages 15+. Total deaths in each category of cause are also reported by sex as 'all'.

dth (integer)
Annual number of deaths for each category of age and sex, by cause.

class (string)
The highest level of classification in the systems used in the Annual Reports. The classification systems used varied over the period. Before 1901 hierarchical classification schemes were employed that included class and order descriptors (for example, individual causes were classified in the order 'miasmatic diseases' within the class 'specified febrile or zymotic diseases'). After 1900 causes of death were grouped into numbered higher order categories (designated by the variable classno), but no names were assigned to these categories.

classno (integer)
Number assigned to highest level of classification used in the Annual Reports.

order (string)
Secondary level of classification in Annual Reports, nested within class.

orderno (integer)
Number assigned to each order, in the Annual Reports.

causeno (integer)
'Cause of death number'. Causes of death were not numbered individually before 1911 except in the period 1877-1882, when causes were assigned numbers within each order. From 1911-1949 each cause was numbered sequentially. Note that the number assigned to each cause varied between coding schemes, even where the cause itself retained the same descriptor.

Sublevel (string)
Lowest level of numerical descriptor. Before 1911 alphabetic codes were used occasionally to subdivide an order-level category. After 1910 alphabetic codes were used to distinguish between causes grouped together under a single cause of death number (causeno), for example in 1911 the cause 'smallpox' was given the number 5, but the cause was further subdivided into 5a: 'smallpox – unvaccinated'; 5b: smallpox – vaccinated'; and 5c: 'smallpox – vaccination doubtful'.

5. Checks and corrections.
The causes of death are transcribed as they appeared in the Annual Reports. Numerical checks were performed to check for accuracy of data entry during entry, using row and column totals. Some corrections have been made to the numbers of deaths by cause where there were inconsistencies within the printed tables – this usually required the correction of totals to match the sum of deaths by individual age groups.

Please report any errors you find to Romola Davenport (rjd23@cam.ac.uk)

6. References

Annual Reports of the Registrar-General of Scotland, 1855-1949, HMSO, Edinburgh (1855-1919 available on HISTPOP: http://www.histpop.org/).

Cameron, Anne (2007) 'The establishment of civil registration in Scotland', *The Historical Journal*, 50(2): 377-395.

Crowther, Anne (2006) 'By death divided. Scottish and English approaches to death certification in the nineteenth century', Paper presented at the Society for the Social History of Medicine conference, Warwick, June 2006 (http://www.gla.ac.uk/media/media_82267_en.pdf)

Davis Gayle (2009) 'Stillbirth registration and perceptions of infant death, 1900-60: the Scottish case in national context', *Economic History Review* 62(3): 629-654.