

## Notes to 'Annual deaths by cause, age and sex in England and Wales, 1848-1900'.

1. Introduction
2. The dataset
3. The variables
4. The classification systems
5. Checks and corrections
6. References

### 1. Introduction: cause of death data for England and Wales in the nineteenth century.

Cause of death data were collected for England and Wales from the introduction of civil registration in 1837<sup>1</sup>. However cause of death data were not published systematically until 1855, from which date they are available in a continuous annual series for England and Wales by age and sex<sup>2</sup>. Cause of death data were published in a variety of forms, with different degrees of temporal and spatial resolution and of aggregation of causes. Deaths were published annually by age, sex and cause at a high level of resolution, usually by single cause, for London, and for England and Wales. The present dataset contains the annual deaths by cause for England and Wales by age and sex. Cause of death data were also published by registration district, both annually without breakdown by age or sex in the Annual Reports, and as decadal averages by age (and sometimes sex) in decennial Supplements to the Annual Reports. However at the registration district level cause of death data were published at a lower level of resolution, with aggregated categories of causes that are not consistent between decades. There are also data available on deaths by cause on a weekly or quarterly time scale for some administrative areas, in the annual reports of the Medical Officers of Health for those areas, although (to my knowledge) without breakdown by age or sex<sup>3</sup>. There are therefore various trade-offs between geographical resolution and resolution at the temporal, sex and cause-specific levels. The decadal cause of death data by registration district have been made machine-readable by Robert Woods and are available in the AHDS archive (SN3552 and SN4570<sup>4</sup>). The present dataset is complementary to the registration district dataset.

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 allow 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 have been used previously to follow individual causes of death, but the absence of a fully

---

<sup>1</sup> See Hardy (1994) for an account of the development of cause of death registration in England and Wales, and Williams (1996) for a description of the registration and classification process.

<sup>2</sup> Published in the Annual Reports of the Registrar-General of Births, Deaths, and Marriages in England [and Wales]. The Annual Report for 1855 contains a summary table of annual deaths by cause, age and sex for the years 1848-1854. The causes of death are presented in more aggregated form than in subsequent years, but are still more detailed than the categories used for registration district data.

<sup>3</sup> See Woods (1984) for an analysis of such data for the Borough of Birmingham in the late nineteenth century.

<sup>4</sup> The data for 1851-1900 were deposited by Woods (SN3552), and subsequently the data for 1851-1910, digitized by Woods, were deposited by H.R. Southall (SN4570, file `mort_cause_0` – note that the dataset is misdescribed with respect to period in the accompanying documentation). These decennial registration district-level data for cause of death by age are analysed in Woods & Shelton (1997) and Woods (2000).

- 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, and for 5 year age groups up to age 14 or 24 (depending on the year). 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 migration between smaller units makes it impossible to follow birth cohorts, whereas migration is less important at the national level.
  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.

### **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, such as the registration district data described above, where possible.
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. Some causes are not presented singly but aggregated (for example, 'Prematurity and Debility' constituted a single category before 1858, when Debility was transferred to a new category of 'Atrophy and Debility'. Therefore none of these causes can be followed singly over the whole period).
4. 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 may be some under-recording of infant deaths, and some misrecording of ages, although under-recording is not considered to be a problem by 1880<sup>5</sup>). In addition the data are highly aggregated and lack any information about cause-specific mortality by occupation group, income etc. The decennial Supplements to the Annual Reports do report male adult death rates for some years by occupation from at least 1860.

### **2. The dataset.**

The dataset contains annual cause-specific mortality data by age and sex for England and Wales for the years 1848-1900, transcribed from the Annual Reports of the Registrar-General of Births, Deaths, and Marriages in England and Wales. The dataset consists of a single file spanning the period 1848-1900 ('dth\_1848-1900.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 classification systems were assembled and coded as separate Excel files and then merged with the death data in STATA and checked for consistency. The dataset is available as a tab-delimited text file ('dth\_1848-1900.txt') that can be imported into Excel, Access or statistical software packages. The dataset structure is simple. Each

---

<sup>5</sup> See Woods (2000) chap. 2 for a discussion of these issues.

line contains a single cause of death for a particular year, age group and sex. There are 11 variables, described in section 4.

### 3. The classification systems.

There were a number of changes in the nosology employed in the Annual Reports over the period 1848-1900. In addition there were a number of changes in descriptors and numbers of causes over the period that meant that it was not possible to employ a single uniform coding system in the dataset. The coding systems in the dataset are therefore divided into seven periods, described below. The classification systems used in the Annual Reports varied over the period but always employed a hierarchical scheme of class, order and specific cause of death. Causes are coded uniformly within each of the seven periods, with the exception of 'Diphtheria' which has different codes in 1858 and 1859.

1. **1848-1854.** The data for 1848-1854 are derived from a summary table in the Annual Report of 1855 and present deaths assigned to 107 major causes in 6 classes (minor causes were aggregated with related major causes and cannot be distinguished). Between 1848 and 1880 deaths were classified into five named classes, with a residual group of causes that included ill-defined and missing causes, and from 1860, 'Abscess or Tumour'. Therefore a 6<sup>th</sup> class, 'Miscellaneous', was introduced in the dataset to cover causes not included in the five classes recognised in the Annual Reports. Age classes for 1848-1854 are 0, 1, 2, 3, 4, 5-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85-94, 95+, and 'unknown'.
2. **1855-1857.** From 1855 the data were published annually, and from 1855-1874 deaths by cause were published annually in a main table containing the major causes of death, and in a supplementary table of minor causes of death. Deaths due to minor causes were also included in the main table, but were aggregated with deaths due to related major causes, without mention of the minor causes. In the electronic dataset all causes of death from both main and supplementary tables are presented, with deaths from major causes disaggregated and correctly distributed between major and minor causes. The code for each cause of death consists of a string of 4 numbers (e.g. 1.1.10.0), where the final number indicates that the cause was either listed in the main table (last number is 0), or was a minor cause aggregated in the main table (last number is other than 0). Deaths were assigned to 263 causes, in 6 classes (the number of causes in any year may be less than 263, because causes were not recorded if no deaths occurred under that heading). The age groups are 0, 1, 2, 3, 4, 5-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85-94, 95+.
3. **1858-1859.** Deaths were assigned to 263 causes, in 6 classes, with some reorganisation and refinement of the previous scheme. For example, Class 1 (Zymotic) was reorganised into 4 orders and expanded to incorporate causes such as 'Want of Breast-milk', 'Privation' and 'Intemperance' previously classified under 'Violent Causes'. Class 4 was renamed from 'Diseases of Growth, Nutrition, and Decay' to 'Developmental Diseases', with a new classification of diseases by age ('Children', 'Adults' and 'Old People', as well as 'Diseases of Nutrition'). *Users should be aware that the coding system for 1858-59 is not consistent with the previous coding scheme, and therefore the same causes may have different codes in 1855-57 compared with 1858-59 (likewise, the same code may apply to different causes over the two periods). Note also that Diphtheria has different codes in 1858 and 1859.* The age classes are as for 1855-1857 for 1858, but from 1859 the age group 15-24 was divided into 15-19 and 20-24.
4. **1860-1874.** Deaths were assigned to 330 causes in 6 classes, representing an expansion of the previous scheme. Minor causes were still reported in separate supplementary tables

- and have been treated as described in 2 above. Age classes are 0, 1, 2, 3, 4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85-94, 95+.
5. **1875-1880.** From 1875 all causes were presented in a single table (instead of main and supplementary tables). The same classification scheme was used as for 1858-1874 (but codes may be different due to the reordering that accompanied the integration of minor with major causes). Deaths were assigned to 283 causes in 6 classes. Age classes are as for 1860-1874.
  6. **1881-1890.** A new classification scheme was introduced, with eight classes. The main change was to class 1 (Zymotic diseases), which was redefined to exclude 'Parasitic' and 'Dietetic' diseases, which then formed separate classes. The classification of 'Constitutional' and 'Developmental' diseases was simplified, and that of 'Local' Diseases expanded. Several diseases, such as 'Dropsy', 'Atrophy' 'Debility' previously classified into other classes were now classified as 'Ill-defined or not Specified' (this class also included causes classified under 'Miscellaneous' pre-1881 in the dataset). From 1880 all causes appeared every year, regardless of whether deaths occurred in these years. Exceptions to this are 'Battle' and 'Plague' (note that military deaths and deaths on the seas/overseas were excluded from the calculations of deaths for England and Wales, but can be found in separate tables in the Annual Reports). There were some changes in nomenclature over the period, which have been preserved in the dataset. For example the residual deaths in each order were classified under 'Other Diseases of...[e.g. Circulatory System]' pre-1886, but under 'Other and undefined Diseases of ...[e.g. Circulatory System]' from 1886. Where the change was clearly only nominal then the same code has been applied to both causes. Age classes are 0, 1, 2, 3, 4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ (i.e. the oldest age group is 85+ rather than 95+ as it was pre-1881). From 1888 onwards the age group under one year old was subdivided into 0-2 months, 3-5 months and 6-11 months, with a separate total for deaths under age one.
  7. **1891-1900.** The causes are classified according to the scheme introduced in 1881, but with some changes to the orders 'Diseases of the Lymphatic System' and 'Accidents and Negligence', and the reduction of the order 'Suicide' to a single undifferentiated cause. The numbering systems therefore do not correspond exactly between 1881-1890 and 1891-1900. Two further changes were made in 1899: 'Gastro-enteritis' was added as a new cause of death (6.5.9), and the cause 'Other ill-defined and not specified Causes' (8.1.8) was divided into two new causes: 'Other ill-defined Causes' (8.1.8.1) and 'Not Specified Causes' (8.1.8.2). Age classes were as for 1888+ (i.e. the age group under one year old was subdivided into 0-2 months, 3-5 months and 6-11 months, and the last open interval was 85+).

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\_1848-1900.txt' contains 11 variables, described below.

***cod (string)***

Causes of death. The Annual Reports contain a fairly comprehensive list of causes by age and sex for England and Wales. This is in contrast to the smaller number of aggregated categories in which deaths by cause were reported for registration districts annually and in the decennial Supplements to the Annual Reports. However there was still some aggregation of minor with major causes and also of accidental deaths (although finer breakdowns of accidental causes were presented periodically as separate tables in the Annual Reports). This is discussed in section 3.

The terminology and classification of causes of death is not uniform over the period, and the various classificatory schemes are described in section 3. Note that before 1881 the number of causes listed varies by sex and from year to year, because causes with no deaths were not listed in that year. After 1880 the full classificatory system was presented each year, although there were slight changes between years (described in sections 3.6 and 3.7) that mean that the total numbers of causes differ slightly for some years. *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).*

***year (integer)***

Calendar year of death.

***sex (string)***

Male ("M") or female ("F").

***age (string)***

The age groupings are not completely consistent over the period. Deaths were recorded for single year age groups for ages 0-4 for every year 1848-1900 (note that age 0 is labelled as "under1"). In addition deaths under one were broken down into ages 0-2 months, 3-5 and 6-11 months for the years 1888-1900. Deaths were recorded in five year age groups for ages 5-14 for all years, and also for ages 15-24 for years 1859-1900. Above age 24 (or age 14 for years 1848-1858) data were presented in ten year age groups up to the last open interval, which was 95+ for years 1848-1880 and 85+ for 1881-1900. The dataset contains all age groups as reported in the Annual Reports. Totals are presented for under 1 (1888-1900), under 5 (1848-1900) and all ages (1848-1900). Age unknown ('unknown') exists as a value for the years 1848-1854.

***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 in the Annual Reports varied over the period but always employed a hierarchical scheme of class, order and specific cause of death. Classes and orders were named and numbered, for example Class 1 was designated "Zymotic" (before 1881), or "Specific Febrile or Zymotic Diseases" (1881-1900).

***classno (integer)***

Number assigned to each class 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.

***codno (integer)***

‘Cause of death number’. Causes were numbered sequentially within each Order in the Annual Reports. However these numbers are not reproduced in the dataset because some years include deaths from Supplementary tables, which had a different numbering system, and because numbering was not consistent across years within each coding period in the Annual Reports. Instead a number was assigned to each specific cause of death within an Order in the electronic dataset in such a way that each number was consistently associated with a particular cause within each of the coding periods described in section 3, except where the classification changed slightly during a coding period (see the next variable).

***code (string)***

Numeric string assigned to each individual cause of death in the electronic dataset, incorporating the class number, order number and cause of death number (e.g. 2.1.1 is the cause numbered 1 in class 2, order 1 of that year). The code variable was included to make it simpler to specify particular causes in a query, by using the code instead of the full name of the cause. Unfortunately the code is a text string, so can’t be used like a numerical system to specify a range of causes using < and > functions. For causes that are grouped within the classification system it is fastest to use the year and classno or some combination of year, classno, orderno and codno to specify the members of the group. For years before 1880 each cause has a four number code, the last digit specifying whether the cause was designated as a major (code ending in ‘0’) or minor (code ending in a number>0) cause of death (see section 3.2 below). Where an individual cause of death has several variants of spelling or punctuation the same code is used for all variants. However the code for a particular cause is not constant throughout the dataset but changes with the classification scheme. In some cases (especially for the years 1855-1859) the same code may describe different causes, due to differences in the original coding scheme in the Annual Reports. In some cases causes were shifted from one major category to another (e.g. Lupus was moved from the ‘Skin’ to the ‘Dropsy’ category after 1858), with a consequent change in the code. Therefore users must check the code AND cod for each year when constructing time series.

5. Checks and corrections.

The causes of death are generally named as they appeared in the Annual Reports, but some changes have been made to keep spelling consistent. For instance ‘Whooping-cough’ also appears in the Annual Reports as ‘Hooping Cough’ and ‘Whooping Cough’, but has been standardised to ‘Whooping-cough’. In other cases small variations in punctuation have been preserved, although the code is the same for all variants.

Numerical checks were performed to check for accuracy of data entry. Totals of deaths within each class and order (column totals) and total deaths for all ages for each cause (row totals) were published in the Annual Reports. These were compared with totals in the electronic dataset and the electronic version corrected where necessary.

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.

6. References

*Annual Reports of the Registrars-General of Births, Deaths, and Marriages in England [and Wales], 1855-1900*, HMSO, London.

Hardy, A (1994) Death is the cure of all diseases. *Social History of Medicine*, 7: 472-492

Williams, N (1996) The reporting and classification of causes of death in mid-nineteenth-century England. The example of Sheffield. *Historical Methods*, 29: 58-71.

Woods, R (1984) Mortality and sanitary conditions in late nineteenth century Birmingham. In Woods, R & Woodward, J (eds.) *Urban Disease and Mortality in Nineteenth Century England*, Batsford, London: 176-202.

Woods, R (2000) *The Demography of Victorian England and Wales*. Cambridge University Press, Cambridge.

Woods R & Shelton N (1997) *An Atlas of Victorian Mortality*. Liverpool University Press, Liverpool.