

Fluctuations in and possible unreliability of death statistics for small sub-groups of the population in small areas

Vital Events statistics are produced from complete counts of all the events which were registered, and so are not subject to some of the kinds of errors that may affect the results of sample surveys. However, the figures for a small sub-group of the population, particularly in a small area, may be subject to large percentage fluctuations from (say) year to year. It follows that the figure for a small sub-group of the population in a small area in any given year may provide an unreliable indicator of the usual annual number of events. This note illustrates this point using the numbers of deaths at ages 59 and under for individual datazones for 2001 to 2007 (the years for which the National Records of Scotland (NRS), formerly the General Register Office for Scotland (GROS), conducted the ad-hoc analysis whose results are described below; a separate page explains what datazones are). A workbook containing the tables is available via a link at the foot of the page. The first table is Table 6 because the analysis of datazones' numbers of deaths at ages 59 and under followed the analysis of their numbers of deaths of all ages, which is described in the [Fluctuations in numbers of deaths for datazones](#) section. Some of the points on this page refer to some of the tables which are available from that page.

Deaths of people aged 59 and under for each datazone for each year

The numbers of deaths of people aged 59 and under in individual datazones were analysed for two reasons. First, as an example of a population sub-group for which the numbers per datazone are very small. Second, because a user of NRS statistics suggested that the proportion of deaths of people aged 59 and under "however crude and arbitrary that may seem, would still be valuable as an indicator of poverty, deprivation and general health status ... because it's not based on sampling or estimates, it shouldn't be troubled by the problems of the confidence interval" - but it would **not** be a good indicator for small areas (such as datazones), for reasons which are given later.

In 2007, there were 8,051 deaths of people aged 59 or under, an average per datazone of 1.2. [Table 6](#) shows that:

- 34% of datazones had no deaths of people aged 59 or under;
- 33% had just one such death;
- 18% had two such deaths;
- 8% had three such deaths; and
- 6% had four or more such deaths.

Areas with (e.g.) hostels for people with particular types of problem may tend to have higher numbers.

When the data for 2001 to 2007 were used to calculate the (rounded) average number of deaths at ages 59 and under for each datazone, a slightly different pattern was found (see [Table 7](#)):

- 12% of datazones had a (rounded) average of 0 deaths per year of people aged 59 or under;
- 58% of datazones had a (rounded) average of 1 such death per year;
- 23% had a (rounded) average of 2 such deaths per year;
- 6% had a (rounded) average of 3 such deaths per year; and

Vital Events – Deaths – Background Information

- only 1% had a (rounded) average of 4 or more such deaths per year.

[Table 8](#) and [Table 9](#) show the numbers of deaths of people aged 59 and under in the datazones which had the largest numbers of deaths of all ages ([Table 8](#)) and the smallest (non-zero) numbers of deaths of all ages ([Table 9](#)). This is for consistency with the lists of the same datazones which appear in [Tables 3 and 4](#) (which are available via the page on [Fluctuations in numbers of deaths for datazones](#)). Therefore, [Table 8](#) does not show the 20 datazones which have the largest numbers of deaths of people aged 59 and under, and [Table 9](#) does not show the 20 datazones with the smallest (non-zero).

[Table 10](#) gives figures for 2001 to 2007 for the datazones which provided 'randomly selected' examples of each (rounded) value of the annual average number of deaths of all ages (ranging from 0 to 36) - the same datazones as appear in [Table 5](#) (which is available via the section on [Fluctuations in numbers of deaths for datazones](#)).

[Table 10](#) shows that these datazones' numbers of deaths aged 59 or under fluctuated greatly (in percentage terms) from year to year. For example:

- the datazone with a rounded average of 3 deaths (of all ages) per year had an annual average of 0.4 deaths aged 59 or under, but it had three such deaths in 2005 and none in any of the other six years;
- the one with a rounded average of 6 deaths (of all ages) per year had an annual average of 1.0 deaths aged 59 or under, but three of its seven deaths were in 2004, and all seven were in the first four years of the period;
- the one with a rounded annual average of 10 deaths (all ages) averaged 0.4 deaths aged 59 or under, but two of its three deaths were in one year (2002).

In addition to the normal year-to-year fluctuation in their numbers of natural deaths, some datazones' figures for some years will be affected by factors such as tragic events (e.g. a bad car accident).

The proportion of deaths at age 59 or under would not be a good indicator for datazones, because both its numerator and the denominator would usually be small and could be subject to large percentage year-to-year fluctuations, leading to potentially large percentage fluctuations in the proportion. Because the denominator would be small, the proportion would not have many possible values (e.g. for a datazone with 4 deaths, the only possible values for the indicator would be 0, 0.25, 0.5, 0.75, and 1). So, just one death aged 59 or under could produce a result which was well above the overall figure for Scotland (0.144), and the series of values for a given datazone could fluctuate markedly from year to year (e.g., for the datazone with a (rounded) average of 3 deaths (all ages) per year, the proportion would have been 0.375 in 2005 and zero in each of the other six years).

Because datazones' numbers of deaths are generally small, even combining the data for (say) five years to try to 'smooth out' the fluctuations in the figures will not help much - the overall average of 8.6 deaths (of all ages) per datazone per year is an average of 43 per five-year period. From statistical theory, with an underlying rate of 43 per five-year period, natural variation could well produce values of, say, 30 and 55 in different five-year periods. An example of such variability appears in [Table 5](#) (which lists the 20 datazones with the most deaths, of all ages, which is available via the section on [Fluctuations in numbers of deaths for datazones](#)) the ninth datazone had an average of 43.9 deaths per year, and its number of deaths in the individual

Vital Events – Deaths – Background Information

years varied between 30 and 59 (the full series is: 42, 53, 30, 49, 39, 59 and 35 - so there is very little to suggest any change in that area's underlying rate of death, and it appears that its year-to-year changes are due to 'random' variability).

Even using data for 5-year periods, the proportion of deaths at age 59 or under could be a very volatile indicator for an individual datazone because the denominator (its total numbers of deaths in 5 years) could fluctuate a lot in percentage terms, and the numerator (its total numbers of deaths aged 59 and under), being much smaller, could fluctuate much more in percentage terms. The suggested proportion would also be difficult to interpret, as it could be affected by tragic events. There is no point in producing it when the Scottish Index of Multiple Deprivation and its various components provide many deprivation-related indicators for individual datazones.

The workbook containing Tables 6 to 10 is available at the following link:

[Deaths of people aged 59 and under for each datazone for each year](#) (Excel 48 Kb)