Standardisation by industry of fatal injury rates by country and region, 2010/11 to 2014/15p

Contents

Summary

The number and incidence rates of worker fatalities appear to vary, when comparing one GB region within another. Taken at face value, some regions have a higher or lower number of deaths per head of worker population than others – for example as published in the annual fatal injury statistics release. However, simple comparisons like this can present difficulties in making conclusions about whether one region is 'more or less risky' than another. Two issues are prominent, and addressed in this analysis. Firstly is whether the industrial composition of each region has an effect; and secondly whether a low number of deaths in a given region could produce potentially misleading results.

Analysis shows that the small number of deaths in individual regions for a single year does have an adverse effect on interpreting analysis; conversely looking at deaths for the latest five years combined improves the reliability of the analysis.

Based on the latest five years but before standardising for industry, there are three regions where the fatality rates are statistically higher than GB: Wales; Scotland; and Yorkshire and the Humber. There are also three regions where the fatality rates are statistically lower than GB: south east; north east; and London.

After standardising by industry, rates for some regions have effectively reduced, others increased. Yorkshire and the Humber is the only region that remains statistically higher than GB, albeit by a small margin. There are two regions that remain statistically lower than GB, namely London; and south east. Wales, Scotland, and the north east are no longer statistically different than GB.

Introduction

Taken at face value, the number of annual workplace deaths* does vary region-by-region, and even when differing employment levels between regions are taken into account, the corresponding injury rates also vary between regions. One argument for the differences in rates between regions is that the type of work is inherently more (or less) risky in one region compared to another.

The European Commission also recognise this argument, with Eurostat (the statistical arm of the EC) producing standardised rates of fatal and non-fatal injury for all member states. Eurostat standardise injury data by 'industry', on the basis that some industries are inherently more risky than others. For consistency, the following analysis of GB fatality rates uses the same methodology as Eurostat, to standardise by 'region' of GB (regions as defined by the Office for National Statistics - ONS).

When the question 'are there any differences between regional fatality rates?' is used, we need to (i) define this statistically, as to whether there is a 'statistically significant' difference between the

**There is a further subtle question to consider: whether we are measuring the number of deaths, or the number of fatal incidents, such that one incident could feasibly lead to more than one death. The following analysis takes the former, on the basis that multiple fatalities are quite un-common, although these incidents may need to be taken into account when understanding some of the statistics.*

rates. And (ii) understand whether any standardisation changes the picture compared to un-standardised rates.

It is likely that relatively few users of statistics may be aware that fatality rates based on actual counts are subject to 'error'. Here the term 'error' is not meant to be a mis-calculation or human error, but random error that occurs naturally. This random error may be substantial when the counts of deaths (numerator), and/or counts of employment (denominator), are small, e.g. less than 15 deaths. A fatality rate observed in a single year can be considered as a sample or estimate of the true or underlying rate. This idea of an underlying rate is an abstract concept, since the rate in one year did actually occur. However as annual rates do fluctuate each year, sometimes dramatically, it is the underlying rate that policy interpretation should seek to address. Hence the larger the numerator of the observed rate, the better the observed rate will estimate the underlying rate.

In a single year many regions may have only a few deaths, and such rates may fluctuate greatly from year to year. One means of addressing this problem is to look at five-year rates, as in this analysis, where the numerator and denominator will be larger.

Methods - overview

The following analysis uses fatal injury worker numbers in Great Britain, for the period 2010/11 to 2014/15 (provisional), where the enforcing authority is either HSE or local authority. Data where the region of incident is not known is excluded (normally no more than one or two per year, usually railways–related or offshore). Definitions of 'region' are defined by ONS. Employment estimates used to calculate fatality rates is sourced from ONS, via the Annul Population Survey (APS) jobs series. Rates of injury are expressed as the number of workers killed, per 100,000 employed workers.

Regional data on the number of deaths in a single year can be very small, for example just one death was recorded in the North East in 2014/15. Statistically it is not possible to analyse such small numbers, and even some regions with a higher number of deaths gives a wide range of uncertainty in some of the results. To reduce this variability, the main analysis was performed on the latest five years combined, 2010/11 to 2014/15p. This time period was chosen as a compromise: long enough to provide sufficiently robust estimates (to minimise the confidence intervals); and short/recent enough to use data that is relevant to current working conditions, i.e. not 'dated'.

Eurostat standardise injury rates for each member state (compared to the whole EU rates), by industry, namely SIC 2007 'section (letter)' level, and the technical method used is 'direct' standardisation. The following also uses the same standardisation process, and compares each GB region in turn with the whole GB fatality rate over the latest five years.

Analysis

Taking the latest single year (14/15 provisional) of published worker fatal injury numbers and rates, Table 1 and Figure 1 show each region in comparison to GB. The rates figures are all per 100,000 workers (APS jobs). Whilst injury rates for each region as shown in Table 1 are higher or lower than GB, it is the variability of those rates in Figure 1 that makes comparisons difficult. The variability increases dependent upon one or two factors: the number of deaths in a region; and the level of employment in a region – the smaller the number(s), the larger the variability. Table 1

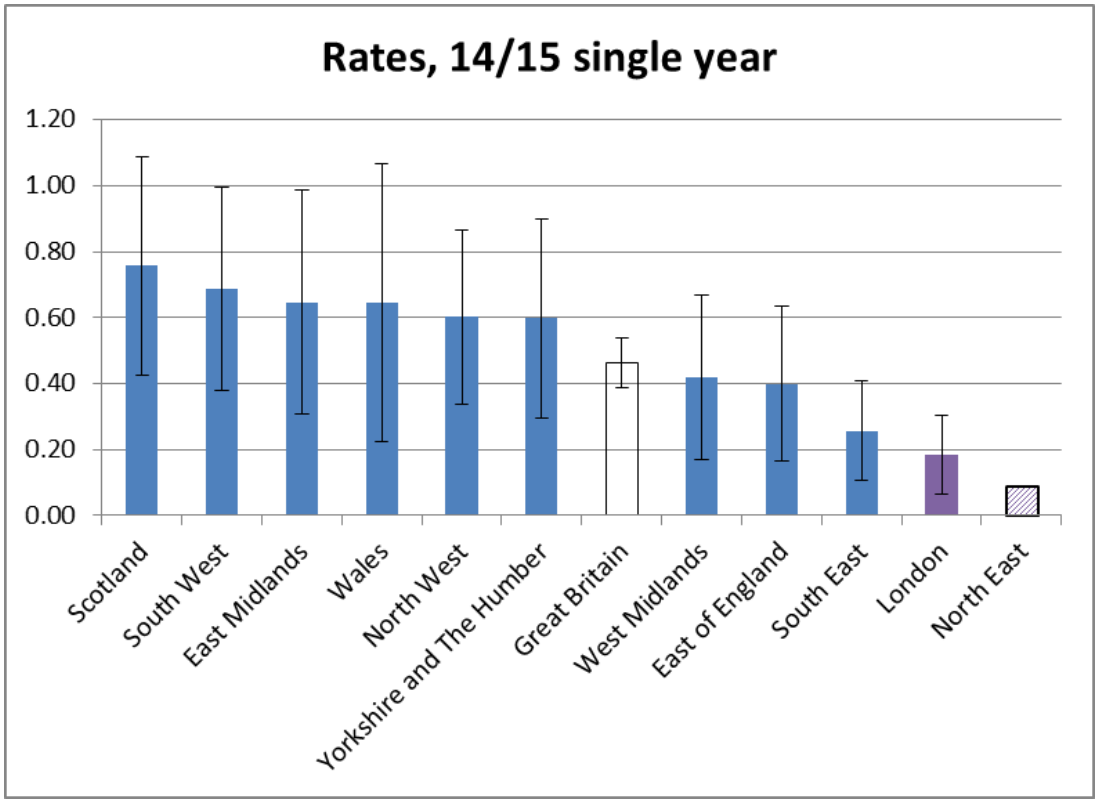
also shows whether each region in turn is statistically significantly different than GB (based on a 95% confidence interval).

Based on a single year, there are two regions statistically different than GB, namely London; and the south east (both being lower than GB). Comparisons with the north east have not been calculated, as there is only one recorded fatality.

Table 1: Number and rates of worker fatalities, 14/15p, un-standardised

	14/15		Sig diff
Country/Region	**Number**	**Rate**	**from GB?**
Scotland	20	0.76	No
Wales	9	0.65	No
North East	1	0.09	N/A
North West	20	0.60	No
Yorkshire and The Humber	15	0.60	No
East Midlands	14	0.65	No
West Midlands	11	0.42	No
East of England	11	0.40	No
London	9	0.18	Yes
South East	11	0.26	Yes
South West	19	0.69	No
Great Britain	**142**	**0.46**	

Figure 1: Rates of worker fatalities, 14/15p, un-standardised



Statistical practice suggests there should be a lower limit in the numerator (the number of deaths), below which results become less reliable. As a statistical rule of thumb, anything less than around

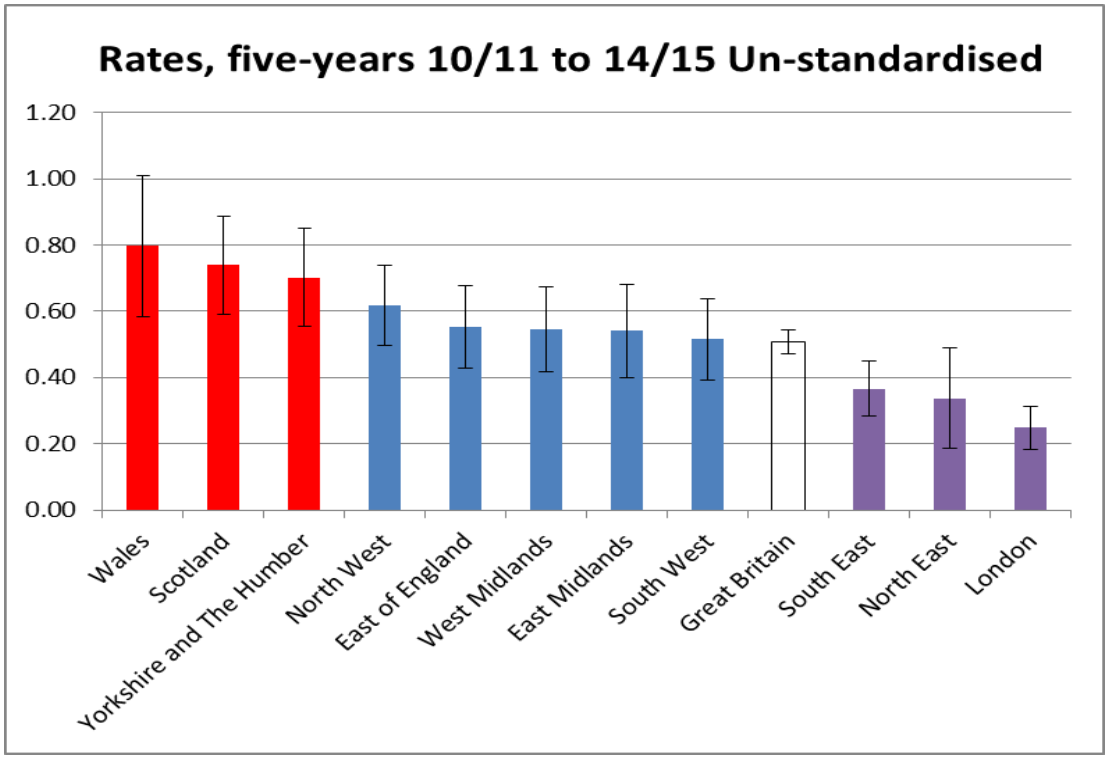
15 in the numerator produces a confidence interval of the same magnitude as the estimate itself. For example in Figure 1 above, in Yorkshire and the Humber there were 15 deaths, the rate is 0.6, and the confidence interval is also 0.6, giving the rate a 95% CI of 0.3 to 0.9. This figure of 15 could have been in the range as low as 8 or as high as 25, simply due to chance.

Another rule of thumb suggests avoiding numbers where the relative standard error (RSE) is more than a certain amount, often 30%. This would give a lower limit of around 11 deaths per single region, below which results may be less reliable. Table 1 for the single year 14/15 suggests a number of the regions are near or below these lower limits, hence reliable analysis may be difficult. It is possible to increase the reliability (by reducing the variability), and Table 2 and Figure 2 look at the deaths for the latest five years combined, 2010/11 to 2014/15p.

Table 2: Number and rates of worker fatalities, 10/11 to 14/15p, un-standardised

	10/11 to 14/15 Un-standardised		Sig diff
Country/Region	Number	Rate	from GB?
Scotland	96	0.74	Yes
Wales	54	0.80	Yes
North East	19	0.34	Yes
North West	102	0.62	No
Yorkshire and The Humber	86	0.70	Yes
East Midlands	57	0.54	No
West Midlands	69	0.54	No
East of England	75	0.55	No
London	57	0.25	Yes
South East	77	0.37	Yes
South West	69	0.52	No
Great Britain	**761**	0.51	

Figure 2: Rates of worker fatalities, 10/11 to 14/15p, un-standardised



The variability for each region has been reduced, as indicated by the smaller error bars in Figure 2. As the number of deaths per region has increased, all are now higher than the lower threshold of around 15, as the north east has the lowest deaths at 19.

Figure 2 indicates that:

- There are three regions where the fatality rates are statistically higher than GB: Wales; Scotland; and Yorkshire and the Humber.
- There are three regions where the fatality rates are statistically lower than GB: south east; north east; and London.
- The remaining regions do not appear to be statistically different from GB.

The five-year rates in Table 2 are 'un-standardised', in that no account is taken for the distribution of deaths, and employment, according to the industry of work. For example if one region had more agriculture (and fewer service-sector) workers than another, we would 'expect' to see more deaths in that region, simply as agriculture overall has a higher fatality rate than service industries.

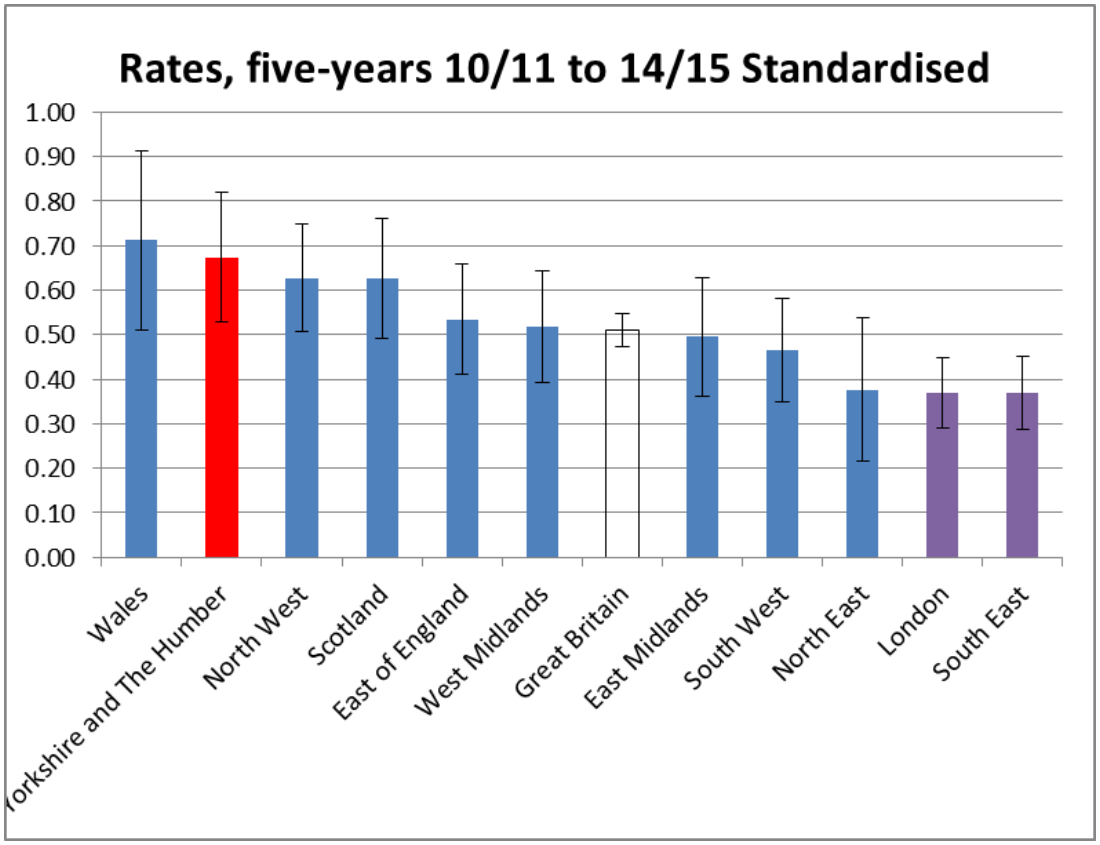
Hence to standardise the rate for each region, a weight is calculated for each industry section, and calculated as: the 'section' employment for GB, as a proportion of all GB employment. This weight is then applied to the region rate for the same section. For example in agriculture, this weighting will have the effect of reducing the contribution of the agriculture fatality to the overall rate for a region, where that region has a higher proportion of agriculture workers than in GB.

Table 3 and Figure 3 both show the effects of standardising by industry:

Table 3: Number and rates of worker fatalities, 10/11 to 14/15p, standardised

	10/11 to 14/15 Standardised		Sig diff
Country/Region	Number	Rate	from GB?
Scotland	96	0.63	No
Wales	54	0.71	No
North East	19	0.38	No
North West	102	0.63	No
Yorkshire and The Humber	86	0.67	Yes
East Midlands	57	0.49	No
West Midlands	69	0.52	No
East of England	75	0.53	No
London	57	0.37	Yes
South East	77	0.37	Yes
South West	69	0.46	No
Great Britain	**761**	0.51	

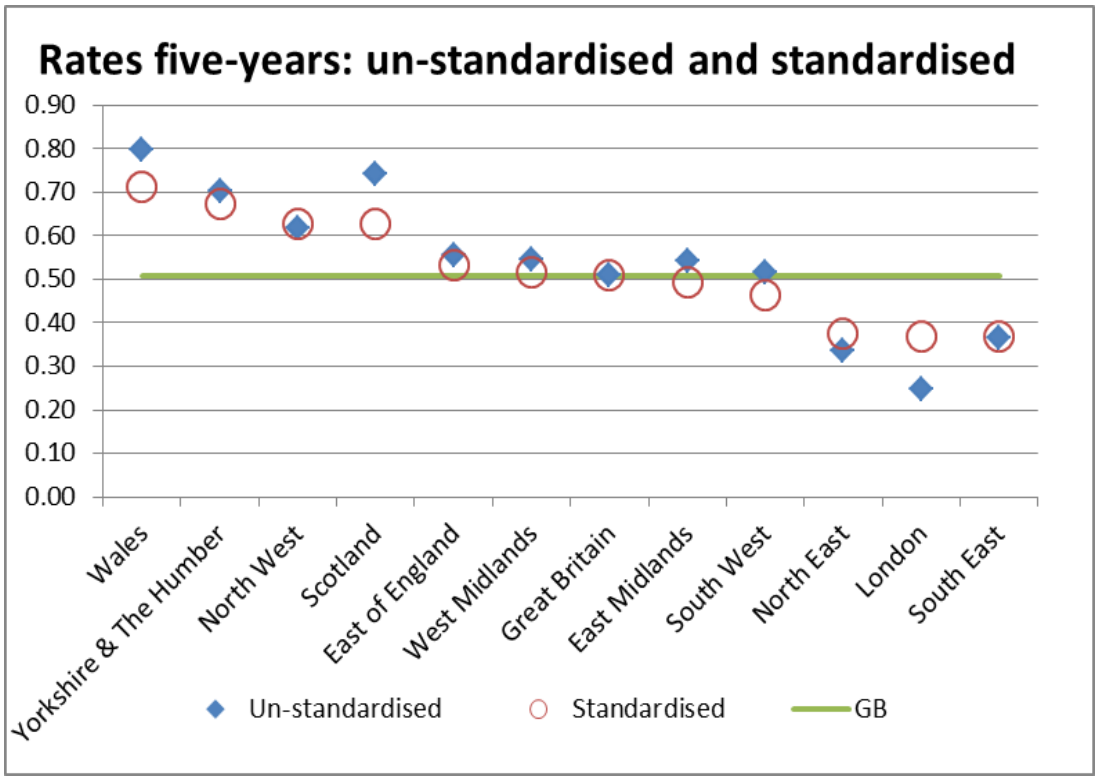
Figure 3: Rates of worker fatalities, 10/11 to 14/15p, standardised



The effect of this standardisation, shown in Figure 4 below (without confidence intervals for clarity), is:

- There is one region that is statistically higher than GB, namely Yorkshire and the Humber, albeit by a small margin;
- There are two regions that are statistically lower than GB, namely London; and south east;
- Other regions that prior to standardisation were statistically higher or lower, are no longer as such (Wales; Scotland; and north east).

Figure 4: Effect on rates of standardisation, five-year averages



Interpreting the results and sensitivity analysis.

- Yorkshire and the Humber, after standardisation, remains statistically higher than GB. However, Y&H has a rate that is lower than Wales; but noting Wales is not statistically higher than GB. This difference of result is because the employment levels of Y&H are nearly double that of Wales, which translates into a wider confidence interval for Wales.
- Similarly the north east is not statistically lower than GB, whereas London, and the south east, are statistically lower – despite all three regions having very similar fatality rates. This again is because employment in the north east is substantially lower than the other two regions, again resulting in wider confidence intervals for the north east.
- Looking again at Y&H, there were 86 deaths in the five-year period. If this figure was lowered by two deaths (to 84), the difference in rates would no longer be statistically different.
- In contrast to Y&H, the number of deaths in London, and the south east, would both have to increase by a greater extent (at least 10 extra deaths per region), to make these regions no longer statistically significantly lower than GB. Any reasons as to why these two regions remain significantly lower than GB are beyond the scope of this analysis.
- In Wales there were two fatal accidents, with four deaths in each incident (i.e. a total of eight deaths from two incidents), and by coincidence both incidents took place in the same year (2011). This pattern of multiple deaths is unusual, normally there are one or two multiple death incidents per year across all GB, with usually two or maybe three deaths per incident. The statistical principles used here don't properly apply in these cases, as there is an assumption that all events (deaths) are independent of each other. To maintain statistical principles, if the eight deaths are replaced by two (one death per incident), this reduces the five-year rate by around 10 per cent from 0.80 to 0.71 (and from 0.71 to 0.65 after standardising). Neither of these reduced rates are statistically different from GB. The conclusion on this point, is that multiple deaths do have an impact on the figures when making comparisons.

Technical Annex.

Analysis is based on assumptions of a Poisson distribution of deaths. One of these assumptions is that events are independent, which can present some problems when one incident results in more than one death. As these events are relatively infrequent, it does not normally present too much of an issue.

Calculations.

(i)

Confidence intervals of individual counts are calculated using the inverse gamma distribution.

(ii)

The standard error and 95% confidence interval of an individual regional or GB fatality rate are given by:

$$SE=\sqrt{\frac{r(1-r)}{n}}$$

$$95\%\ CI = r \pm 1.96 \times \sqrt{\frac{r(1-r)}{n}}$$

where r is the fatality rate, and n is the employment.

(iii)

To test if two rates are statistically significantly different from each other:

$$SE=\sqrt{\frac{r1(1-r1)}{n1}+\frac{r2(1-r2)}{n2}}$$

Statistically significantly different (at 95% level) if difference between rates > 1.96 x SE

(iv)

The Eurostat method on standardising the incidence rate is given by:

++++++++++++

The method for calculating standardised incidence rates for countries is the following:

1. The first step is to calculate for each year a fixed set of EU weights wk for each economic sector k of the common branches A, C – N. The EU weights wk differ only for sectors and thus are based on EU sector totals only (the same for each country). They correspond simply to the ratio between the EU reference population of a sector (number of all workers in this sector in the EU) and the total EU reference population for all sectors (all EU workers):

$$w_k=\frac{refpopk}{refpopEU}$$

where refpopk is the EU (EU-28) reference population for sector k and refpopEU is the entire reference population of the EU common branches A, C – N;

2. Then, for each country X and sector k the weights wk are multiplied with the (non-standardised) incidence rates r in order to produce weighted incidence rates r' specific for each sector in country X:

$$r'_{X,k}=r_{X,k}* \text{ wk}$$

3. Finally, the standardised incidence rate sr of the country X is calculated as the sum of the

weighted incidence rates r' of the country X:

$$sr_x = \sum_{k=A,C}^{N} r'_{X,k}$$

with k representing sectors A, C, D, E, ... until N. Standardised incidence rates have the same unit as the non-standardised ones (number of accidents per 100,000 workers).

For a fuller Eurostat explanation and worked examples, see
http://ec.europa.eu/eurostat/cache/metadata/Annexes/hsw_acc_work_esms_an3.pdf
and
http://ec.europa.eu/eurostat/cache/metadata/Annexes/hsw_acc7_work_esms_an3.pdf

++++++++++++

Other references on direct standardisation:

http://www.healthknowledge.org.uk/e-learning/epidemiology/specialists/standardisation

Data as used in Figures 1, 2 and 3:

GOR	Rate 14/15 single year	1.96*SE of rate
Scotland	0.76	0.33
South West	0.69	0.31
East Midlands	0.65	0.34
Wales	0.65	0.42
North West	0.60	0.26
Yorkshire and The Humber	0.60	0.30
Great Britain	**0.46**	**0.08**
West Midlands	0.42	0.25
East of England	0.40	0.24
South East	0.26	0.15
London	0.18	0.12
North East	0.09	

GOR	Rate five-years 10/11 to 14/15 Un-standardised	1.96*SE of rate
Wales	0.80	0.21
Scotland	0.74	0.15
Yorkshire and The Humber	0.70	0.15
North West	0.62	0.12
East of England	0.55	0.13
West Midlands	0.54	0.13
East Midlands	0.54	0.14
South West	0.52	0.12
Great Britain	**0.51**	**0.04**
South East	0.37	0.08
North East	0.34	0.15
London	0.25	0.06

GOR	Rates, five-years 10/11 to 14/15 Standardised	1.96*SE of rate
Wales	0.71	0.20
Yorkshire and The Humber	0.67	0.15
North West	0.63	0.12
Scotland	0.63	0.14
East of England	0.53	0.12
West Midlands	0.52	0.13
Great Britain	**0.51**	**0.04**
East Midlands	0.49	0.13
South West	0.46	0.12
North East	0.38	0.16
London	0.37	0.08
South East	0.37	0.08

A detailed account of the latest fatal injury statistics can be found at http://www.hse.gov.uk/statistics/fatals.htm

Details of how the figures are used for statistical purposes can be found at www.hse.gov.uk/statistics/sources.htm

For information regarding the quality guidelines used for statistics within HSE see www.hse.gov.uk/statistics/about/quality-guidelines.htm

A revisions policy and log can be seen at www.hse.gov.uk/statistics/about/revisions/

Additional data tables can be found at www.hse.gov.uk/statistics/tables/.

Statistician: David Leigh

Contact: david.leigh@hse.gsi.gov.uk

Last updated: October 2015

Next update: October 2016