# Technical Briefing 3

**ASSOCIATION OF PUBLIC HEALTH OBSERVATORIES**

## Commonly used public health statistics and their confidence intervals

## Purpose

This is the third in a series of technical briefings produced by the Association of Public Health Observatories (APHO), designed to support public health practitioners and analysts and to promote the use of public health intelligence in decision making.

In this briefing we describe the calculation of the more common types of statistic used within public health intelligence, specifically: rates, proportions, means and age-standardised rates and ratios. We also describe the use of confidence intervals, i.e. what they are and why they are used, and present the APHO-preferred methods for calculating the intervals for the types of statistic described. Further material, including detailed bibliography, methods and tools, is available from http://www.apho.org.uk

## Contents

## Author

Daniel Eayres
(National Centre for Health Outcomes Development)

## Acknowledgements

# Key public health measures

## Counts, proportions, percentages and rates

The most basic measure used in public health is the count. This may be a count of events such as deaths or admissions to hospital, or a count of people with a particular attribute such as people who smoke. This count itself is essential information for planning the health services for prevention and/or treatment. However, to properly investigate the distribution of disease and risk factors and to make comparisons between different populations, the denominator population or population-years at risk in which the count was observed must also be taken into consideration. The simplest way of doing this is to divide the numerator count by the denominator population to give a proportion, percentage or rate.

At this point it is worth distinguishing between these terms. The term 'rate' in particular tends to be rather loosely applied to describe many public health statistics, some of which would be better described as proportions. The distinction is not important to the calculation of the statistic itself but is necessary to best determine the confidence interval.

Proportions are statistics where the denominator is the count of a 'closed' population, and the numerator is the count of members of this population that have a specified characteristic. Common examples in public health occur where a population is evaluated at a finite time point for a certain characteristic, e.g. survey prevalence of smoking or obesity, day cases as a proportion of elective admissions, discharge to usual place of residence, or stillbirths as a proportion of all live and still births.

For other statistics the population is an 'open' one evaluated over a period of time. Individuals may enter or leave the population during this period (through ageing, migration, birth, death, loss to follow-up etc), each contributing different periods at risk. The denominator is the sum of the population-periods at risk experienced by the individuals (or an approximation to this, such as a mid-year population estimate). The numerator is usually a count of events that occurred in the population over the period. We shall refer to such statistics as 'rates' and examples include mortality rates and cancer incidence rates.

Both proportions and rates are frequently multiplied by a scaling factor for presentation purposes, e.g. per 100,000. When this factor is 100 the statistic is usually described as a percentage. In most, but not all, cases percentages are proportions.

## Means

Some public health indicators are reported as the mean of the individual values observed in a sample or population. For example, the mean number of decayed, missing or filled teeth in children, or the mean length of stay in hospital.

## Age-standardised ratios, rates and proportions

Disease and mortality rates may vary widely by age. Such variation complicates any comparisons made between two populations that have different age structures. For example, consider two areas A and B with equal-sized populations and identical crude all-age death rates. At first glance they appear to have a similar mortality experience. Suppose, however, that area A has a younger age structure than area B. Given that mortality rates increase with age, one would expect the older population in area B to experience more deaths. The fact that the two have identical rates means that the younger population in area A must have a relatively worse mortality experience.

The most comprehensive way of comparing the disease experience of two populations is to present and compare their age-specific rates. However, when the number of populations being compared increases, the volume of data that needs to be considered quickly becomes unmanageable. What is needed is a single, easily interpreted, summary figure for each population that is adjusted to take into account its age structure. Such summary figures are calculated using age standardisation methods. It may also be desirable to standardise for other variables, such as sex or level of deprivation, that may also potentially confound any comparisons.

The two most common methods of age standardisation are:

- Indirect: The age-specific rates of a chosen standard population (usually the relevant national or regional population) are applied to the age structure of the subject population to give an expected number of events. The observed number of events is then compared to that expected and is usually expressed as a ratio (observed/expected). A common example is the standardised mortality ratio (SMR).

- Direct: The age-specific rates of the subject population are applied to the age structure of the standard population. This gives the overall rate that would have occurred in the subject population if it had the standard age-profile.

Table 1 shows the advantages and disadvantages of each method.

*Table 1. Advantages and disadvantages of indirect and direct age standardisation.*

| Indirect | Direct |
|---|---|
| **Advantages** | **Disadvantages** |
| • The indirect method requires only the total number of observed events in the subject population and may therefore be used in some situations where the direct method cannot. | • The direct method requires that the observed events in the subject population are available broken down by age. If this information is not available the directly standardised rate cannot be calculated. |
| • Indirect standardisation is more stable as it minimises the variance, giving a smaller standard error and narrower confidence intervals than the direct method. It is therefore more appropriate when dealing with the statistical significance of small populations.[1] | • For small subject populations the age-specific rates of the subject population are based on small numbers and consequently are unstable. Small changes in the number of deaths in a particular age band may result in large changes in the directly standardised rate. |
| **Disadvantages** | **Advantages** |
| • Indirect standardisation does not necessarily preserve consistency between the populations being compared, and in extreme situations may give misleading results. See Figure 1. | • The direct method preserves consistency between the populations being compared, i.e. if each age-specific rate in area A is greater than each of the corresponding age-specific rates in area B, then the directly standardised rate for area A will always be higher than that of area B irrespective of the standard population used.[1;3] Consequently it is the preferred method for comparing a number of different populations against each other. |
| • Indirectly standardised ratios for areas A and B may be compared to the standard but should only be directly compared to each other if the age structures of areas A and B are similar to the standard, or if the ratio of their age-specific mortality rates to that of the standard is consistent across the age groups.[2] | |
| • Indirect standardisation can be used for comparisons over time, but only by freezing the standard rates at a fixed point in time. This is often felt to be less than intuitive. The validity of the comparisons is subject to the same conditions as described above for comparison between areas. | • Directly standardised rates can readily be compared over time provided the same standard population is used. |

Figure 1 illustrates the circumstances in which indirect standardisation can lead to loss of consistency and produce misleading results.
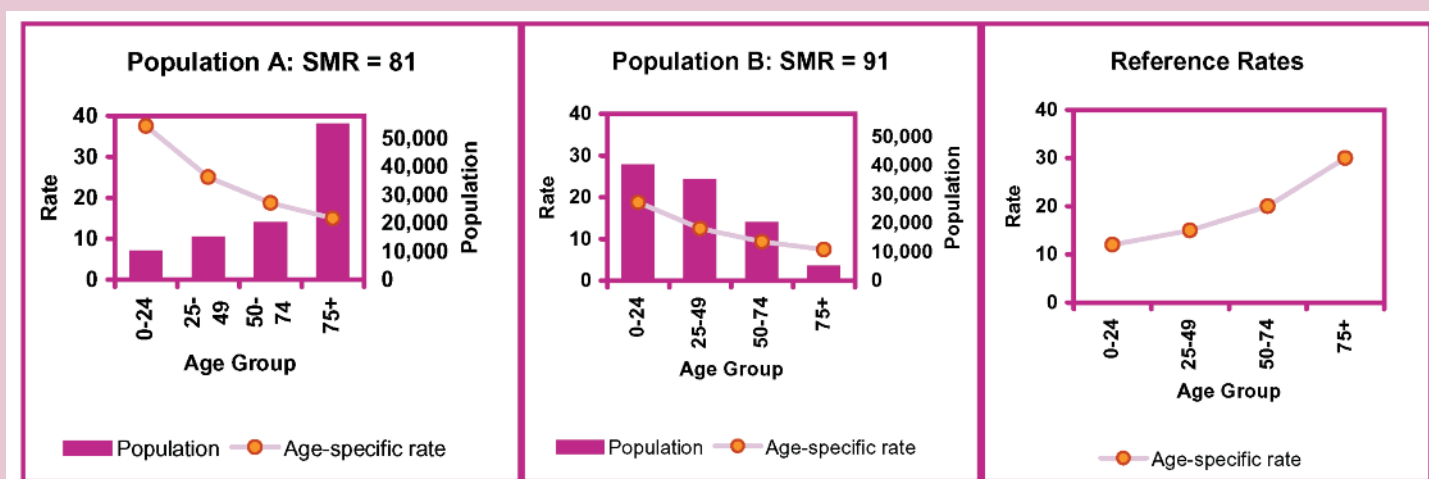


*Figure 1. Example of misleading indirectly age-standardised ratios.*

In Figure 1, even though Population A has double the rate of Population B in every age-group, the SMR for Population A is lower than that for Population B. This arises because the population distribution is very different, and the ratios of the local rates to the reference rates are very uneven. Although it is not unusual for different populations to have very different age-distributions, it is unusual for the ratios to be so different - in general rates will tend to vary with age in a consistent way across different populations.

In practice, the two methods generally give comparable results. Moreover, it has been demonstrated by Breslow and Day[4] that when the two do differ it is not necessarily true that the direct method is the more 'correct'. Where the ratio between the age-specific rates of areas A and B varies by age group the choice of standard population/ rates becomes crucially important for both indirect and direct methods. This choice is often as, or more, important than the choice of method.[5;6]

The European Standard Population (ESP) is often used for direct standardisation. This is a hypothetical population structure which does not change and is the same for both genders. The advantage of using such a hypothetical population is one of greater comparability, for example, between different countries, across time periods, and between genders. However, the ESP dates from the 1970s, and compared to current UK age structures it puts much greater weight on the younger age groups. This can have some undesirable effects as ideally the standard population should be similar to the populations being standardised. For the same reasons it is usually inappropriate to use the ESP for indicators where the denominator is a particular subgroup of the population, for example those with a certain condition or occupation, patients admitted to hospital, etc. The age structure of these subgroups may differ substantially from that of the ESP. In summary, the ESP must be used when comparisons are to be made with other rates which have been standardised to the ESP, but for self-contained comparative analyses, a more suitable (e.g. average) population may be preferred.

## Life expectancy

Life expectancy is another common public health statistic that allows comparisons (of mortality experience) between areas with different population structures. It is based on a life table approach rather than direct or indirect standardisation. The methodology is beyond the scope of this briefing and the reader is referred to other PHO and ONS resources.[7-9]

# Confidence intervals

A confidence interval is a range of values that is used to quantify the imprecision in the estimate of a particular value. Specifically it quantifies the imprecision that results from random variation in the estimation of the value; it does not include imprecision resulting from systematic error (bias).

In many studies the source of this random variation is sampling. Even in the best designed studies there will be random differences between the particular sample group selected and the overall target population of inference. Any measurement taken from the sample group therefore provides an imprecise estimate of the true population value.

In public health many indicators are based on what can be considered to be complete data sets and not samples, e.g. mortality rates based on death registers. In these instances the imprecision arises not as a result of sampling variation but of 'natural' variation. The indicator is considered to be the outcome of a stochastic process, i.e. one which can be influenced by the random occurrences that are inherent in the world around us. In such instances the value actually observed is only one of the set that could occur under the same circumstances. Generally in public health, it is the underlying circumstances or process that is of interest and the actual value observed gives only an imprecise estimate of this 'underlying risk'.

The width of the confidence interval depends on three things:

1) The sample size from which the estimate is derived (or population size if derived from a complete data set). Larger samples give more precise estimates with smaller confidence intervals.
2) The degree of variability in the phenomenon being measured. Fortunately, observed phenomena often are known, or assumed, to follow certain probability distributions, such as the Poisson or Binomial. This allows us to express the amount of variability mathematically, and build it into the confidence interval formulae.
3) The required level of confidence - this is an arbitrary value set by the analyst giving the desired probability that the interval includes the true value. In medicine and public health the conventional practice is to use 95% confidence but it is not uncommon to see alternatives. Within the APHO community 99.8% confidence intervals are increasingly being used alongside 95% intervals to reflect the control limits used in Statistical Process Control approaches.[10] Increasing the level of confidence results in wider limits.

For a given level of confidence, the wider the confidence interval, the greater the uncertainty in the estimate.

Figure 2 demonstrates confidence intervals for smoking prevalence generated by 20 random samples of 100 persons. For sample one the smoking prevalence observed is 16.9% and the associated 95% confidence interval is 10.8% to 25.4%. This means that we are 95% confident that the range 10.8% to 25.4% includes the true population prevalence. On average, 95 out of every 100 of these intervals will include the population value. This can be seen in the figure where 19 of the 20 intervals (ie. 95%) include the 25% population prevalence, the exception being sample 16. In many public health applications we do not have the luxury of repeated samples, but the confidence intervals are calculated in such a way as to ensure that if we did, then 95% of them would contain the true value.
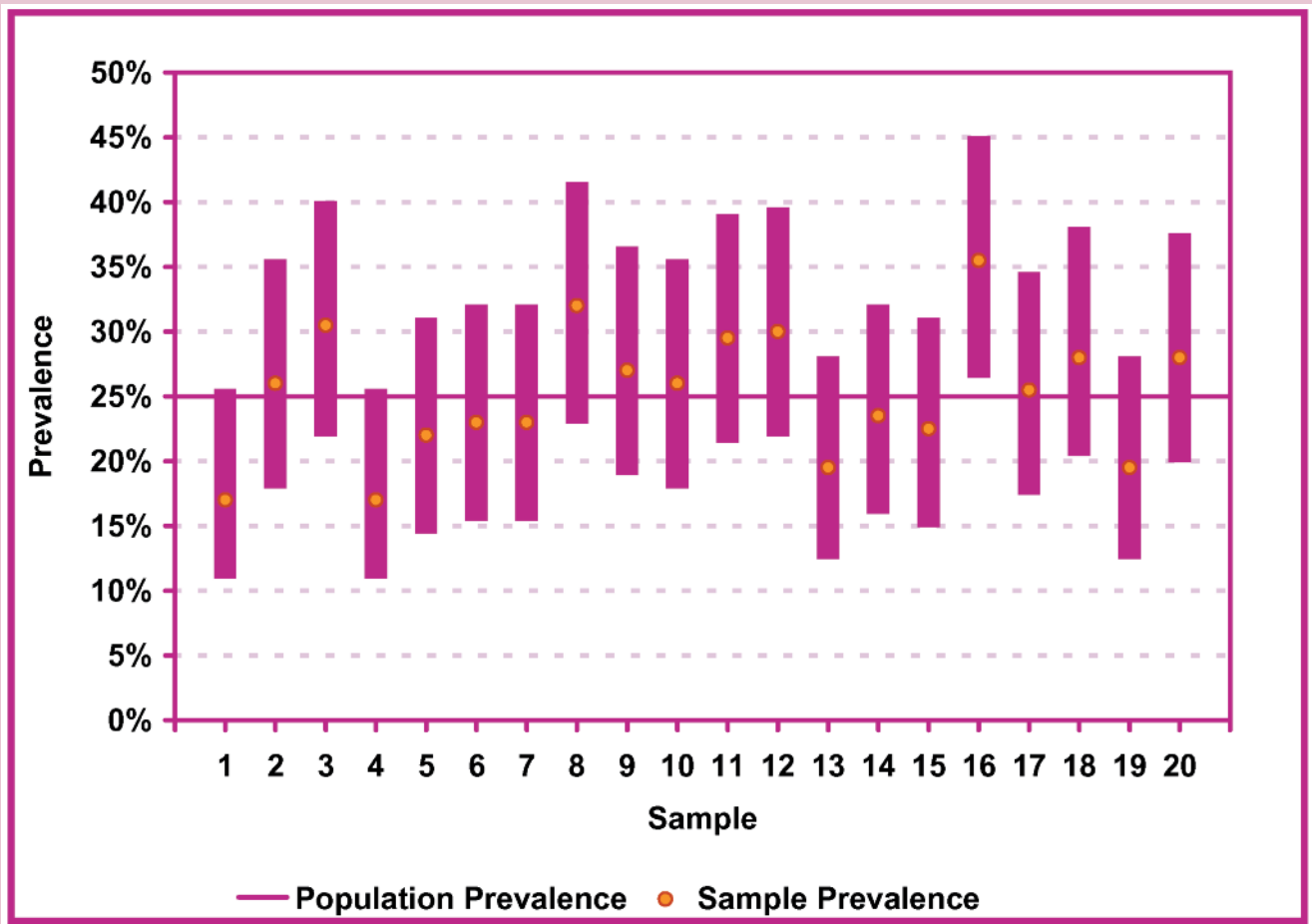
*Figure 2. Smoking prevalence: sample estimates and 95% confidence intervals for 20 random samples of 100 persons from a population with a smoking prevalence of 25%.*

## When should confidence intervals be used?

Confidence intervals should be presented alongside the point estimate wherever an inference is being made from a sample to a population or from a set of observations to the underlying process (or 'risk') that generated them.

## Presentation of confidence intervals

When presenting the confidence interval the level of confidence should be clearly stated and the lower and upper limits of the interval clearly labelled or presented as a range. The use of the $\pm$ sign or a dash to indicate the range should be avoided.

For example:

| Area | Statistic | 95% Confidence Interval | |
|------|-----------|-------------|-------------|
| | | **Lower Limit** | **Upper Limit** |
| Area 1 | 95.2 | 93.5 | 96.9 |
| Area 1 | 95.2 | 93.5 to 96.9 | |
| Area 1 | 95.2 | (93.5, 96.9) | |

For Area 1 the statistic was 95.2 (95% confidence interval: 93.5 to 96.9).

In graphs or charts, the confidence limits are usually represented as bars or whiskers extending above and below the value of the estimate. Figure 2 is a typical example.

## Using confidence intervals for making comparisons

The confidence interval may be used to compare an estimate against a target or benchmark value in a similar way to performing an exact statistical test. In such a test we proceed on the assumption that the true value is equal to the target or benchmark value. The appropriate probability distribution is then used to determine the probability (P-value) of observing a value as extreme, in either direction (a two-sided test), as the actual observed estimate. If the P-value is below a given significance threshold, $\alpha$, then the original assumption is rejected and the difference is said to be statistically significant.

If the target or benchmark value is outside the confidence interval it can be inferred that the difference between the estimate and the target is statistically significant at the corresponding significance level. For instance, if the target value is outside the 95% confidence interval then the difference is significant at the $\alpha = 5\%$ level (P < 0.05).
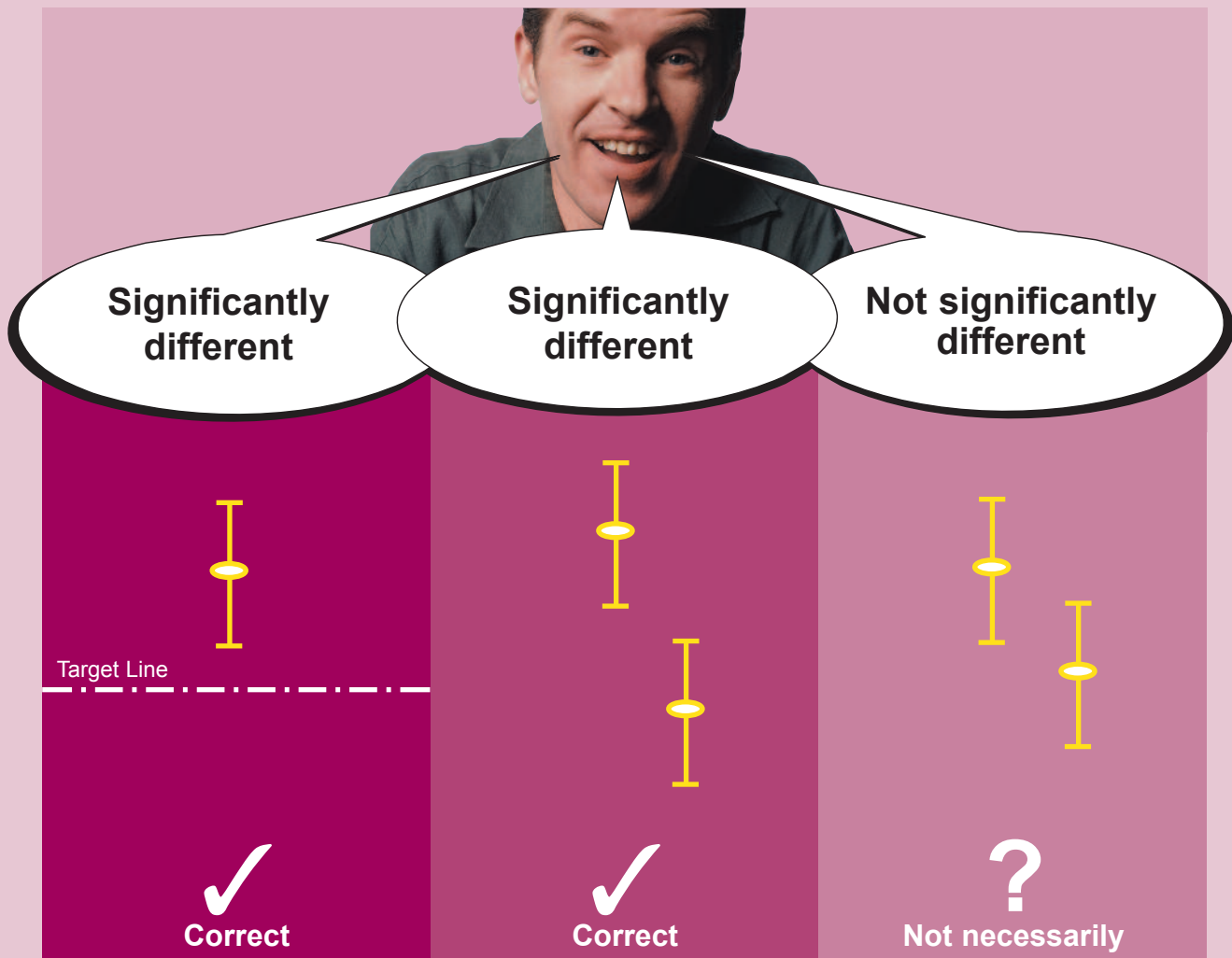
*Figure 3. Using confidence intervals for making comparisons.*

Comparisons are often made between two or more different estimates, for example between different organisations or time periods (see Figure 3). Sometimes in such cases statistical testing is undertaken by comparing the confidence intervals of the estimates to see if they overlap - with non-overlapping confidence intervals being considered as statistically significantly different. This is quick and easy to do, but not necessarily conclusive. Whilst it is safe to assume that non-overlapping confidence intervals indicate a statistically significant difference, it is not always the case that overlapping confidence intervals do not. A more exact approach is to calculate the ratio of the two estimates, or the difference between them, and construct a test or confidence interval based on that statistic. Such methods are not covered in this briefing, but can be found in a standard textbook.

# Calculating the statistics and their confidence intervals

The following section presents the calculation of the statistics and their confidence intervals in detail. The formulae become rapidly more complicated, but please do not let that deter you, as APHO will be turning them into user-friendly spreadsheet tools available from http://www.apho.org.uk

## Proportions

The proportion $p$ is given by: $p = \dfrac{O}{n}$ **Formula 1**

where:
$O$ is the observed number of individuals in the sample/population having the specified characteristic;
$n$ is the total number of individuals in the sample/population.

Confidence intervals for proportions are determined using the Binomial distribution. A Normal approximation method is often presented in statistical textbooks but does not perform well when the numerator and/or denominator is small. The preferred APHO method is the Wilson Score method [11] which has been evaluated and recommended by Newcombe and Altman.[12;13] It can be used with any data values and, unlike some methods, it does not fail to give an interval when the numerator count, and therefore the proportion, is zero.[14]

The 100(1−$\alpha$)% confidence interval limits for the proportion $p$ are given by:

$$p_{lower} = \frac{\left(2O + z^2 - z\sqrt{z^2 + 4Oq}\right)}{2\left(n + z^2\right)}$$

**Formula 2a**

$$p_{upper} = \frac{\left(2O + z^2 + z\sqrt{z^2 + 4Oq}\right)}{2\left(n + z^2\right)}$$

**Formula 2b**

where:
$q = (1−p)$ is the proportion without the specified characteristic;
$z$ is the 100(1− $\alpha$/2)th percentile value from the Standard Normal distribution. For example for a 95% confidence interval, $\alpha = 0.05$, and $z = 1.96$ (i.e. the 97.5th percentile value from the Standard Normal distribution).

## Rates

$$\text{The rate of events } r \text{ is given by: } r = \frac{O}{n} \quad \textbf{Formula 3}$$

where:
$O$ is number of observed events;
$n$ is the population-years at risk.

Provided the rate is low and the denominator at risk is large, the variability in the observed count $O$ is described by the Poisson distribution. This can be used to give a confidence interval for $O$ and hence $r$. The preferred method is Byar's approximation as it is computationally simple and gives very accurate approximations to the exact Poisson probabilities even for small counts.[15] However, tables of the exact probabilities are widely available and should be used for the lower limits of observed counts under 5.

Using Byar's method, the 100(1− $\alpha$)% confidence interval limits for the rate $r$ are given by:

$$r_{lower} = \frac{O_{lower}}{n},$$

$$\text{where } O_{lower} = O \times \left(1 - \frac{1}{9O} - \frac{z}{3\sqrt{O}}\right)^3$$

**Formula 4a**

$$r_{upper} = \frac{O_{upper}}{n}, \text{ where } O_{upper}$$

$$= (O+1) \times \left(1 - \frac{1}{9(O+1)} + \frac{z}{3\sqrt{(O+1)}}\right)^3$$

**Formula 4b**

where:
$O$ is the observed count;
$z$ is the 100(1− $\alpha$/2)th percentile value from the Standard Normal distribution.

Where the rate $r$ is not low, the appropriateness of the Poisson distribution is more debatable but it is still generally used.[16] In extreme circumstances, for example looking at high risk groups in small populations, the use of the Poisson distribution may lead to limits beyond sensible bounds.

## Counts

The limits of the observed count should be determined using Formulae 2a and 2b, or 4a and 4b, depending on whether the count is based on a Binomially distributed proportion or a Poisson rate respectively. For example, for a smoking prevalence indicator the observed count is based on a proportion and its limits should be calculated from the Binomial distribution; for a mortality or cancer incidence count Poisson limits should be used. In both cases the limits for the proportion or rate should be multiplied by $n$ to convert them into limits for the observed count. (In the case of Formula 4a and Formula 4b, this simply means using the elements referred to as '$O_{lower}$' and '$O_{upper}$'.)

## Means

$$\text{The sample mean } \bar{x} \text{ is given by: } \bar{x} = \frac{\sum_{i} x_i}{n}$$

**Formula 5**

Where:
$x_i$ is the observed value for the $i$th individual;
$n$ is the total number of individuals in the sample/population.

When the size of the sample is large (say >50), the sample mean is Normally distributed and its standard error can be estimated by $s/\sqrt{n}$, where $s$ is the observed standard deviation of the individual values in the sample:

$$s = \sqrt{\dfrac{\sum\limits_{i}(x_i - \bar{x})^2}{(n-1)}}$$

The traditional Normal approximation method can be used to give the $100(1-\alpha)$% confidence interval limits:

$$\bar{x}_{lower} = \bar{x} - z\left(\dfrac{s}{\sqrt{n}}\right) \qquad \textbf{Formula 6a}$$

$$\bar{x}_{upper} = \bar{x} + z\left(\dfrac{s}{\sqrt{n}}\right) \qquad \textbf{Formula 6b}$$

where $z$ is the $100(1-\alpha/2)$th percentile value from the Standard Normal Distribution.

When samples are small, $s/\sqrt{n}$ will vary substantially between samples and cannot be considered a reliable approximation to the standard error of the mean. However, if the population distribution of the individual values itself is Normal, with mean $\mu$, then the distribution of $(\bar{x}-\mu)/(s/\sqrt{n})$ follows the Student's t-distribution for $(n-1)$ degrees of freedom. The confidence intervals are then found by substituting the appropriate percentile point values of the Standard Normal distribution with those from the t-distribution with the appropriate degrees of freedom. These values are found from tables or the built-in functions found in most spreadsheet and statistical software.

When using such tools the reader should ensure that they are using the correct set of probabilities (two-tailed) and that they are evaluating the distribution against the appropriate value, i.e. the confidence level or probability level or percentile point, as this requirement may vary between tools.

$$\bar{x}_{lower} = \bar{x} - t\left(\dfrac{s}{\sqrt{n}}\right) \qquad \textbf{Formula 7a}$$

$$\bar{x}_{upper} = \bar{x} + t\left(\dfrac{s}{\sqrt{n}}\right) \qquad \textbf{Formula 7b}$$

where $t$ is the $100(1-\alpha/2)$th percentile value from the t-distribution with $(n-1)$ degrees of freedom. For example for a 95% confidence interval for a mean of 10 observations, $\alpha = 0.05$ and $t = 2.26$ (i.e. the 97.5th percentile value from the t-distribution with 9 degrees of freedom).

For small samples where the individual values are not Normally distributed the data can be adjusted so that they become so by applying a suitable transformation.

## Indirectly age-standardised ratios and rates

The statistic most commonly presented for the indirect method of age standardisation is the standardised ratio - this is the ratio of the observed number of events relative to the number of events that would be expected if standard age-specific rates were applied to the particular observed population's age structure. A common example is the standardised mortality ratio (SMR).

The indirectly standardised ratio (ISR) is given by:

$$ISR = \dfrac{O}{E} \;=\; \dfrac{\sum\limits_{i} O_i}{\sum\limits_{i} E_i} \;=\; \dfrac{\sum\limits_{i} O_i}{\sum\limits_{i} n_i \lambda_i} \qquad \textbf{Formula 8}$$

where:
$O_i$ is the observed number of events in the subject population in age group $i$;
$E_i$ is the expected number of events in the subject population in age group $i$ given the standard rates;
$n_i$ is the number of individuals in the subject population in age group $i$;
$\lambda_i$ is the crude age-specific rate in the standard population in age group $i$.

For presentation purposes, the ratio is usually multiplied by 100. By definition, the standard population will have a ratio of 100. Ratios above 100 indicate that the number of events observed was greater than that expected from the standard rates, and ratios below 100 that it was lower. In some instances the ratio is multiplied by the overall crude rate of the standard population and presented as an indirectly age-standardised rate.

For the purposes of calculating the confidence interval of the ratio, the expected count is considered to be precise. The imprecision in the ratio is therefore dependent only on the imprecision of the observed count.

The $100(1-\alpha)$% confidence interval limits of the ratio are given by:

$$ISR_{lower} = \dfrac{O_{lower}}{E} \qquad \textbf{Formula 9a}$$

$$ISR_{upper} = \dfrac{O_{upper}}{E} \qquad \textbf{Formula 9b}$$

where:
$O_{lower}$ and $O_{upper}$ are as defined in Formula 4a and Formula 4b respectively;
$E$ is the expected count.
The confidence limits as calculated above should then be multiplied by any scaling factor that has been used in presenting the ratio itself.

## Directly age-standardised rates

Directly age-standardised rates express an indicator in terms of the overall rate that would occur in a standard population age-structure if it experienced the age-specific rates of the observed population.

The direct rate (DSR) is given by:

$$DSR = \frac{1}{\sum_i w_i} \times \sum_i \frac{w_i O_i}{n_i} \qquad \textbf{Formula 10}$$

where:
$w_i$ is the number, or proportion, of individuals in the standard population in age group $i$;
$O_i$ is the observed count in the population in age group $i$;
$n_i$ is the observed denominator population-period-at-risk in age group $i$.

The directly standardised rate is a weighted sum of the independent age-specific rates. Therefore, its variance is a weighted sum of the variances of each of those age-specific rates. The preferred method for calculating the confidence interval is one described by Dobson.[17] In this method the exact interval is found for the crude count and then weighted and scaled to give the interval for the directly standardised rate. The weight used is the ratio of the standard error of the DSR to the standard error of the crude count.
For rates that assume the Poisson distribution, the confidence limits for the DSR are given by:

$$DSR_{lower} = DSR + \sqrt{\frac{Var(DSR)}{Var(O)}} \times \left(O_{lower} - O\right)$$

**Formula 11a**

$$DSR_{upper} = DSR + \sqrt{\frac{Var(DSR)}{Var(O)}} \times \left(O_{upper} - O\right)$$

**Formula 11b**

where $O$ is the total number of observed events in the population;
and the limits $O_{lower}$ and $O_{upper}$ of the observed count are determined using Byar's approximation (Formulae 4a and 4b) or exact look-up tables.

The variance of the DSR and the variance of the observed count are estimated by:

$$Var(DSR) = \frac{1}{\left(\sum_i w_i\right)^2} \times \sum_i \frac{w_i^2 O_i}{n_i^2} \quad ; \qquad Var(O) = \sum_i O_i$$

**Formula 11c**

where:
$w_i$ is the number, or proportion, of individuals in the standard population in age group $i$;
$O_i$ is the observed count in the subject population in age group $i$;
$n_i$ is the denominator population-period-at-risk in the subject population in age group $i$.

# References

1.  Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organization; 1987: 72-3.

2.  Tsai SP, Wen CP. A review of methodological issues of the standardised mortality ratio (SMR) in occupational cohort studies. Int J Epidemiol 1986;15(1):8-21.

3.  Inskip H et al. Methods for age-adjustment of rates. Stat Med 1983;2:455-66.

4.  Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organization; 1987: 67.

5.  Goldman DA, Brender JD. Are standardised mortality ratios valid for public health data analysis? Stat Med 2000;19:1081-8.

6.  Higham JH et al. Direct standardisation has no advantage over indirect standardisation. Letter. BMJ 2002; 7 Aug. Available at http://bmj.com/cgi/eletters/325/7357/177#24481 (accessed 18 February 2008).

7.  Flowers J. Life expectancy. INphoRM (3). Cambridge: ERPHO; 2003. Available at http://www.erpho.org.uk/Download/Public/9479/1/3rd %20proof%20INPHORM%20Issue%203.pdf (accessed 18 February 2008).

8.  Williams E et al. Technical report: calculating life expectancy in small areas. Oxford: SEPHO; 2005. Available at http://www.sepho.org.uk/Download/Public/9847/1/Life %20Expectancy%20Nov%2005.pdf (accessed 18 February 2008).

9.  Toson B, Baker A. Life expectancy at birth: methodological options for small populations (National Statistics Methodological Series No 33). London: ONS; 2003. Available at http://www.statistics.gov.uk/downloads/theme_other/G SSMethodology_No_33.pdf (accessed 18 February 2008).

10. Flowers J. Technical Briefing 2: Statistical process control methods in public health intelligence. York: APHO; 2007. Available at http://www.apho.org.uk/resource/item.aspx?RID=3944 5 (accessed 18 February 2008).

11. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927;22:209-12.

12. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998;17:857-72.

13. Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). Statistics with confidence (2nd edn). London: BMJ Books; 2000: 46-8.

14. Agresti A, Coull BA. Approximate is better than 'exact' for interval estimation of binomial proportions. Am Stat 1998;52:119-26.

15. Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organization; 1987: 69.

16. Brillinger DR. The natural variability of vital rates and associated statistics. Biometrics 1986;42:693-734.

17. Dobson A et al. Confidence intervals for weighted sums of Poisson parameters. Stat Med 1991;10:457-62.

## Other contributory material

18. Bland M. An introduction to medical statistics (2nd edn). Oxford: Oxford University Press; 2000.

19. Esteve J et al. Statistical methods in cancer research, volume IV: Descriptive epidemiology. Lyon: International Agency for Research on Cancer, World Health Organization; 1994.

20. Merrick D. Confidence intervals for standardised rates and ratios. Durham: NHS Executive (Northern & Yorkshire Region); 1997.

21. Washington State Department of Health. Guidelines for using confidence intervals for public health assessment. Washington: Washington State Department of Health; 2002.

22. Eayres D. Annex 3 - Explanations of statistical methods used in the Compendium. In Lakhani A, Olearnik H, Eayres D (eds). Compendium of clinical and health indicators. London: The Information Centre for Health and Social Care/National Centre for Health Outcomes Development; 2007.

# Glossary

**Binomial distribution:** the probability distribution that applies to discrete data with only two possibilities - e.g. success or failure, alive or dead. It is used to generate confidence intervals for proportions. Its shape is dependent on two parameters: the number of trials, or observations; and the proportion of successes.

**Normal distribution:** familiar bell-shaped curve, which is a good representation of the probability distribution of many naturally-occurring variables. Its shape is dependent on two parameters: the mean and the standard deviation.

**Parameter:** a quantity that defines certain characteristics of a system, function or distribution.

**Poisson distribution:** probability distribution representing the number of events (e.g. accidents) in a fixed space of time. Its mean and variance are equal, so its shape is determined by a single parameter.

**Probability distribution:** a function that gives the probability of a specific value, or value within a given interval, being observed at random given the underlying parameters.

**Standard deviation/Standard error:** statistical measure of the amount of variation around the mean of a distribution. It is defined as the square root of the variance. By convention, the term 'standard deviation' is used only when the distribution is describing variation in individual values within a population or sample. The standard deviation of a summary measure, such as the sample mean, is known as the 'standard error'.

**Standard Normal distribution:** a Normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. Any Normally distributed variable $x$, with mean $\mu$ and standard deviation $\sigma$, can be transformed into a 'Z-score' which follows the Standard Normal distribution, by using the relationship $z = (x- \mu)/\sigma$. This relationship means that there is no need to publish tables for any Normal distribution other than the Standard Normal distribution.

**Standard population:** a reference population age-structure used for direct age standardisation. It may be a real population, e.g. England mid-year estimates 2006, or a hypothetical one, e.g. European Standard Population.

**Standard rates:** a set of reference age-specific rates used for indirect age standardisation. In nearly all instances these are real rates rather than hypothetical ones and are taken from the reference or benchmark population against which the populations being standardised are to be compared.

**Statistical significance/Statistical test:** to test a given hypothesis, we initially proceed on the assumption that it is true. If, on this basis, an observation as extreme as that obtained would have been very unlikely to arise by chance, we then 'reject' the hypothesis. The cut-off point or 'significance level' below which we consider the probability to be too small is arbitrary, but is traditionally set at 0.05 (5%). Results which lead to the rejection of the hypothesis are said to be 'statistically significant'.

**Transformation:** the application of a mathematical function to a set of data values, e.g. the logarithm, square root or reciprocal. The purpose is to 'transform' the data so that the distribution of values more closely resembles a Normal distribution.

**Variance:** statistical measure of the amount of variation around the mean of a distribution. The greater the variance the more widely distributed are the observed values.

**Z-score:** a value that follows a Standard Normal distribution.

## About the Association of Public Health Observatories (APHO)

The Association of Public Health Observatories (APHO) represents and co-ordinates a network of 12 public health observatories (PHOs) working across the five nations of England, Scotland, Wales, Northern Ireland and the Republic of Ireland.

APHO facilitates joint working across the PHOs to produce information, data and intelligence on people's health and health care for practitioners, policy makers and the public.

APHO is the largest concentration of public health intelligence expertise in the UK and Republic of Ireland, with over 150 public health intelligence professionals.

APHO helps commissioners to ensure that they get the information they need and our websites provide a regular stream of products and tools, training and technical support.

We work with partners to improve the quality and accessibility of the data and intelligence available to decision makers.

We are constantly developing and learning new and better ways of analysing health intelligence data. We use these new methods to improve the quality of our own work, and share them with others.

Updates and more material, including methods and tools to support our technical briefing series, are available through our website at http://www.apho.org.uk

**For further information contact:**
**Association of Public Health Observatories**
Innovation Centre, York Science Park,
Heslington, York YO10 5DG

Telephone: 01904 567658
http://www.apho.org.uk