Age Specific Rates:
One approach to resolving the problem of different population age structures is to compare rates within specific age groups. An age specific rate is a crude rate for an age range, usually 5 or 10 years.

Crude Incidence/Mortality Rates:
A crude incidence/mortality rate is a measure of the number of new cases/deaths due to a disease in a given time span. The crude incidence rate is usually presented as the annual incidence rate and quoted as the number of cases per 100,000 of a defined study population.

Example:
In the South West there were approximately 164 new cases of breast cancer per 100,000 female population in 2000. This crude annual rate was calculated using the following equation:

$$\text{Number of female breast cancers incident in the South West in 2000 (approx 5598)} \times \frac{100,000}{\text{South West female population 2000 (approx 3,411,534)}} = 164.1$$

The crude figure is useful when an indication of the actual burden of cancer on a population is required, for example a hospital manager assessing the resource needs to care for patients in their area. The crude rate is less useful when we wish to observe trends over a period of years or to compare two or more separate geographical regions. This is because the crude rate does not take into account factors such as age. Cancer is predominantly a disease of the elderly and so a population with a higher average age, such as retirement areas, will tend to have a higher crude rate. Direct crude rate comparisons can therefore be misleading. Figure 1 shows that areas with a more elderly population structure such as Dorset have higher bladder cancer crude incidence rates (light blue bars), than areas with less elderly populations such as Gloucestershire.

Age Specific Rates:
One approach to resolving the problem of different population age structures is to compare rates within specific age groups. An age specific rate is a crude rate for an age range, usually 5 or 10 years.

Example:
In the South West there were approximately 318 new cases of breast cancer per 100,000 female population in the 50-54 years age range in 2000. This age specific incidence rate was calculated using the following equation:

$$\text{Number of 50-54 yr female breast cancers incident in the South West in 2000 (approx 709)} \times \frac{100,000}{\text{South West 50-54 yr female population 2000 (approx 223,275)}} = 317.5$$

Figure 2 shows how age specific rates can identify fluctuations in incidence rates within age groups. The increase in the 50-64 age group between 1990 and 2000 is thought to be an effect of the screening programme. Age specific rates are useful when looking at one or two age groupings. Single summary statistics to compare populations over a range of age groups can be generated from age specific rates. These are the age standardised rates and standardised ratios.
Age Standardised Rates (ASR) - Direct Standardisation:

This is the crude rate which would have occurred if the age specific rates of a study population had operated in a reference population known as a ‘standard population’ (see Figure 1 - dark blue bars). The same standard population is used for all rates to be compared. The number in each age group in the standard population is used as a weighting system in the standardisation process. The age standardised rate is calculated as the sum of the crude age specific rates multiplied by the respective proportions represented in the standard population, giving a ‘weighted mean’ (the ASR). In practise it is usual to calculate ASRs with age specific rates in 5 year age bands and European or world population proportions as the standard.

Example:
The age specific rates in a population are 30 per 100,000 in the 0-44 age group, and 150 per 100,000 in the 45+ age group. The standard population being used has 3 million in the 0-44 age group and 2 million in the 45+ age group.
The age specific rates are multiplied by the standard populations to calculate the predicted number of cases in the standard population:
\[\text{i.e. } 30 \times 3000000 = 900 \text{ and } 150 \times 2000000 = 3000 \text{ and } 900 + 3000 = 3900.\]
The age standardised rate is then calculated as
\[\frac{\text{Number of cases (3900)}}{\text{Total standard population (5,000,000)}} \times 100,000 = 78 \text{ cases per 100,000.}\]

Standardised Ratios (SRRs and SMRs) - Indirect Standardisation:

Standardised Registration Ratios (SRRs) and Standardised Mortality Ratios (SMRs) are single widely used indices which are applied when a number of areas are to be compared in relation to a baseline population. For example comparing the incidence of a disease in each of the health authorities in the South West, relative to the South West as a whole. The age specific rates for the baseline population (the South West) are applied to the corresponding age groups in the study population (the HA). This gives the total expected number of cases/deaths for the study population which is then compared with the observed number of cases/deaths in the study population. The resulting ratio is then multiplied by 100 and expressed as a percentage.

Example:
In the South West as a whole the age specific rate is 30 per 100,000 in the 0-44 age group and 180 per 100,000 in the 45+ age group. The population in one health authority (HA) is 30,000 and 40,000 respectively. If the incidence in the HA was the same as in the South West there would be 81 cases:
0-44 yrs: Study Population (i.e. 30,000) / 100,000 x Baseline Population Age Specific Rate 0-44 yrs (i.e. 30) = 9 expected
45+ yrs: Study Population (i.e. 40,000) / 100,000 x Baseline Population Age Specific Rate 45+ yrs (i.e. 180) = 72 expected
90 cases of breast cancer were actually registered in the HA.
The SRR is the ratio of observed to expected, which is then expressed as a percentage: (90/81) x 100 = 111%.

Prevalence:

Prevalence is the number of people currently living with a disease, and provides a useful indication of the number of patients being actively treated for their disease. The prevalence of a cancer is determined by incidence and survival. If a cancer has a high incidence and individuals tend to live for a long period of time before death then the prevalence will be high. A cancer with a high incidence and a high death rate such as lung cancer will have a lower prevalence than a cancer with a high incidence but a better prognosis, such as breast cancer.

5 year prevalence assumes that individuals are disease free if they have survived to five years from diagnosis. It could be calculated at a set date as the number of people who have been diagnosed with cancer in the preceding five years who are still alive. For cancers which may recur many years after the original diagnosis, particularly breast, bladder and paediatric cancers, 10 or 20 year prevalence is a more appropriate measure.

Confidence Intervals:

The statistical values described in this factsheet are frequently presented with upper and lower confidence limits. Confidence limits can be used as a guide to assess whether there is a significant difference between samples. The usual degree of confidence presented is 95% i.e. the statistic presented is 95% certain to fall within the upper and lower confidence limits.

Crude, Relative and Corrected Survival:

There are several methods for calculating survival and many factors which can affect the result. The reference ‘Cancer Registration Principles and Methods’ gives a good basic description of cancer survival analyses.

Factors Affecting Survival

Factors affecting survival may be demographic and include age, sex, severity of the disease, cancer type, period of diagnosis, place of residence and type of treatment. These factors can be investigated using a variety of regression models of which the most widely used is Cox’s Proportional Hazards Model. Other factors are methodological. The choice of cases upon which to perform survival analyses and the precise definition of variables (e.g. date of diagnosis) are important when preparing a dataset. Cases registered solely from information contained on a death certificate are usually excluded for quality reasons.
Crude, Relative and Corrected Survival:

**Crude Survival using the Direct Method**

The direct method calculates the proportion of patients still alive at the end of a specified interval using only the patients in the study population. It is used when limited information about the patient (alive vs. dead) is known. For example, if 42 individuals of a population of 100 breast cancers are known to have died within 5 years of their date of diagnosis, the 5-year crude survival is (58/100) = 58%.

**Actuarial Methods (indirect or life table methods)**

These methods use the follow-up information available such as date of diagnosis and date of death and have the advantage over the direct method as cases lost to follow up can be accommodated. Actuarial methods use uniform time intervals which can be varied. The following example describes the actuarial method of calculating observed crude survival:

**Example**

If 42 individuals of a population of 100 breast cancers are known to have died within 5 years of their date of diagnosis, and 6 are lost to follow up, the 5-year observed crude survival can be calculated as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of deaths</th>
<th>No. lost to follow up</th>
<th>No. alive at end of year</th>
<th>(1) proportion surviving year</th>
<th>(2) proportion surviving from diagnosis to end of year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21</td>
<td>1</td>
<td>78</td>
<td>0.79</td>
<td>0.79*</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>0</td>
<td>71</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>2</td>
<td>62</td>
<td>0.87</td>
<td>0.63</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>57</td>
<td>0.92</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>53</td>
<td>0.93</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>52</td>
<td>0.98</td>
<td>0.53</td>
</tr>
</tbody>
</table>

If we assume that yearly calculations are to be made then the status of each patient is required at annual intervals from the start date. The proportion of patients surviving each year is multiplied to give a cumulative total survival. In the table this is the proportion surviving from diagnosis to end of the first year (0.79*) multiplied by the proportion surviving the second year (0.91) which gives 0.72. The figures in column (2) are all calculated in this way. The overall 5-year crude survival rate for the population is thus 53%.

**Kaplan-Meier Method**

This method is similar to the actuarial method but instead of calculating cumulative survival at the end of the uniform periods of time (i.e., one year in the example) the proportion of patients surviving can be calculated at intervals determined by the accuracy of the dataset.

**Corrected Survival**

The corrected survival takes into account the specific cause of death allowing cases dying from the diagnosed disease itself to be distinguished from cases dying with the disease but from unknown or unrelated causes. The use of this method is not routinely applied due to the bias introduced by inaccuracy of death certificates.

**Relative Survival**

The relative survival rate is the ratio of the observed (crude) survival rate in a patient group to the expected survival rate in a group of people from the general population similar to the patient group with respect to age, sex, and calendar period of observation. Information on the actual cause of death is not required. Expected survival probabilities are calculated from published population life-tables. Relative survival allows comparisons between different areas with different population structures. It is not a measure of clinical prognosis.

Relative survival rate = \( \frac{\text{Observed survival rate}}{\text{Expected survival rate}} \times 100 \)

Further Information: