3 Vital Statistics and Measurement

3.1 Vital Statistics

The basic measures for assessing the health of a community and its needs for health services are the size and composition of the human population and the counts of vital events occurring within them (births, deaths, morbid disorders, etc). Such measures are known as vital statistics.

Indices, that is summary statistics, derived from such raw data in terms of rates, ratios and proportions are of use to many workers in the health field. Understanding these indices and their uses should help towards efficient provision and use of resources and to appropriate preventative measures targeted at susceptible sections of the community. By themselves, the raw data are of little use until they are transformed to standard indices for valid comparisons.

3.2 Population and Demographics

3.2.1 Population Size

There is only one accurate way to estimate the size of the relevant population, and describe its demographic characteristics, and that is to count it at a particular point in time. This is known as a census and is usually done once every ten years. The more recent census in the UK took place in 2001.

3.2.2 Inter-censal events

The second major problem of measurement in public health is to record the demographic events that occur within the community: births, deaths, marriages and migration both in and out of the area. If these are measured accurately and continuously then the changes in population from year to year can be determined, and the frequent need for censuses eliminated.

In the United Kingdom, every birth, marriage and death in the country is registered. From these registers it is possible to calculate the rates at which births and deaths are occurring in different areas, social groups, ages and at different times of the year. Immigration and emigration are measured only for the whole country by sample estimates at ports and airports.

Population size estimates between census years are only an approximation. Some estimates, such as the numbers in different occupations, are so unreliable that mortality data for them is only tabulated in census years.

3.3 Incidence, Prevalence and Rates

3.3.1 Rates

A rate is defined as the number of events, for example deaths or cases of disease, per unit of population, in a particular time span. To calculate a rate we require:

- a defined period of time

- a defined population, with an accurate estimate of the size of the population during the defined period
• the number of events occurring over the period.

\[
\text{rate} = \frac{\text{No. events}}{\text{total person-time at risk}}
\]

For a fixed time period \( \Delta t \), an average size of the population at risk during that period \( \bar{N} \) and the number of events \( A \) the rate is

\[
\text{rate} = \frac{A}{\bar{N} \times \Delta t}
\]

3.3.2 Incidence

The \textit{incidence rate} refers to the number of new cases of a particular disease that develop during a specified time interval.

For a fixed time period \( \Delta t \), a population at risk of size \( N \) and the number of new cases of disease \( A \) the incidence rate is

\[
\text{incidence} = \frac{A}{N \times \Delta t}
\]

An incidence rate lies between 0 and \( \infty \), and a yearly incidence rate measures the number of cases per person-year.

Measuring the population at risk may be difficult because of changes in the population over the time period. Since it may not be possible to measure disease-free periods precisely, we often calculate the incidence using the average size of the population. This is reasonably accurate if the population size is stable and the incidence rate is low.

3.3.3 Prevalence

The \textit{prevalence} refers to the number of cases of disease that exist at a specified point in time. It is the proportion of the population who have a disease at a given time. Prevalences may also be calculated over a time period; for example the number of events within a time period.

\[
\text{prevalence} = \frac{\text{Number of cases}}{\text{Number at risk}}
\]

Diseases with high incidence rates may have low prevalences if they are rapidly fatal.

3.3.4 Crude Mortality Rate

The \textit{crude mortality rate} is usually calculated as deaths per 1000 population per year. Let \( D \) be the number of deaths in a given time period of length \( \Delta t \), and \( \bar{N} \) be the average size of the population at risk during that period (often approximated by the number in the population at the mid-point of the time period). Then the crude mortality rate is given by

\[
r = \frac{d}{\bar{N} \times \Delta t} \times 1000
\]
3.3.5 Specific Rates

Rates may be required for particular sections of a community and these are referred to as specific rates; that is, where the populations are specified; that is the denominators. For example, age-specific or age- and sex-specific rates may be used for comparison of different populations. Other common specific rates are area, occupation or social class specific (and combinations of these).

3.3.6 Other commonly-used rates

- **infant mortality rate**
  Number of deaths under one year of age after live birth, divided by the number of live births

- **neonatal mortality rate**
  Number of deaths at 0 to 27 days after live birth, divided by the number of live births

- **stillbirth rate**
  Number of stillbirths, divided by the total number of births, live and still

- **perinatal mortality rate**
  Number of stillbirths and deaths at days 0 to 6, divided by the total number of births

- **birth rate**
  Number of live births per year, divided by total population

- **fertility rate**
  Number of live births per year, divided by number of women 15-44 (ie of childbearing age)

Table 1: Mortality Data by Age Group for South West of England 2003. *Source: ONS Mortality Statistics Series DH1 no. 36*

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Population (1000s)</th>
<th>% in age group</th>
<th>Deaths</th>
<th>Age-specific mortality rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>259.7</td>
<td>5.2</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>5-14</td>
<td>609.3</td>
<td>12.2</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>15-24</td>
<td>591.0</td>
<td>11.8</td>
<td>251</td>
<td></td>
</tr>
<tr>
<td>25-44</td>
<td>1319.3</td>
<td>26.4</td>
<td>1219</td>
<td></td>
</tr>
<tr>
<td>45-64</td>
<td>1282.2</td>
<td>25.6</td>
<td>5944</td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>470.7</td>
<td>9.4</td>
<td>8668</td>
<td></td>
</tr>
<tr>
<td>75+</td>
<td>467.1</td>
<td>9.3</td>
<td>39480</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4999.3</td>
<td>100</td>
<td>55865</td>
<td></td>
</tr>
</tbody>
</table>
3.4 Standardisation

Table 2: Mortality Data by Age Group for England 2003. Source: ONS Mortality Statistics Series DH1 no. 36

<table>
<thead>
<tr>
<th>Age Group (1000s)</th>
<th>% in age group</th>
<th>Deaths</th>
<th>Age-specific mortality rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>2848.2</td>
<td>5.7</td>
<td>3682</td>
</tr>
<tr>
<td>5-14</td>
<td>6299.8</td>
<td>12.6</td>
<td>725</td>
</tr>
<tr>
<td>15-24</td>
<td>6304.0</td>
<td>12.6</td>
<td>2634</td>
</tr>
<tr>
<td>25-44</td>
<td>14485.7</td>
<td>29.1</td>
<td>14001</td>
</tr>
<tr>
<td>45-64</td>
<td>11971.3</td>
<td>24.0</td>
<td>62755</td>
</tr>
<tr>
<td>65-74</td>
<td>4158.6</td>
<td>8.3</td>
<td>87714</td>
</tr>
<tr>
<td>75+</td>
<td>3788.3</td>
<td>7.6</td>
<td>331898</td>
</tr>
<tr>
<td>All</td>
<td>49855.9</td>
<td>100</td>
<td>503409</td>
</tr>
</tbody>
</table>

When comparing populations, we can eliminate the effects of, for example, different age structures by looking at age-specific rates. However, this can be cumbersome, and it is often easier to compare a single summary figure.

3.4.1 Direct Standardisation

For direct standardisation, we use a standard population structure for reference. We then calculate the overall mortality rate that this reference population would have observed if it had the age-specific mortality rates of the population of interest.

Suppose the reference population has population counts $N'_k; k = 1, \ldots, K$ in each age-group $k$. We calculate the age-specific mortality rates $r_k$ for the population of interest. The directly standardised rate is given by

$$\text{directly standardised rate} = \frac{\sum_{k=1}^{K} N'_k r_k}{\sum_{k=1}^{K} N'_k}$$

3.4.2 Indirect Standardisation

For indirect standardisation, we take the age-specific rates from the reference population and convert them into the mortality rate we would observe if those reference rates were true for the age-structure of the population of interest. This gives us the expected rate for the population of interest, if age-specific mortality rates were the same as for the reference population.

We calculate the age-specific mortality rates $r'_k$ for the reference population. Suppose the population of interest has population counts $N_k; k = 1, \ldots, K$ in each age-group $k$. The expected rate of deaths in the population of interest is

$$\text{expected rate} = \frac{\sum_{k=1}^{K} N_k r'_k}{\sum_{k=1}^{K} N_k}$$

and the expected number of deaths in the population of interest is

$$E = \sum_{k=1}^{K} N_k r'_k$$
3.4.3 Standardised Mortality Ratio

We can compare the expected number of deaths, using the indirect standardisation method, with the observed number using the standardised mortality ratio (SMR). Let $O$ be the observed number of deaths in the population of interest, and $E$ be the expected number of deaths when indirectly standardised with respect to some reference population.

$$\text{SMR} = \frac{O}{E} \times 100$$

The SMR is a ratio, not a rate or a percentage. An SMR of 100 means that the population of interest has the same number of deaths as we would expect from the reference population. If it is greater than 100, we have more deaths than expected; if it is less than 100 we have less.

3.4.4 Confidence Intervals for SMRs

Suppose that deaths are independent of one another, and occur randomly in time so we can use the Poisson distribution with unknown parameter $\lambda$:

$$O \sim \text{Poisson}(\lambda)$$

with

$$E[O] = \lambda \quad \text{var}(O) = \lambda$$

and so we can approximate $\text{var}(O) \approx O$. Since the expected number of deaths is calculated from a large sample we can treat it as a constant (since its variance is small enough to be negligible) so

$$E[\text{SMR}] = E \left[ \frac{100 \times O}{E} \right] = \frac{100\lambda}{E}$$

$$\text{var(\text{SMR})} = \text{var} \left( \frac{100 \times O}{E} \right) = \frac{100^2 \lambda}{E^2}$$

Provided the number of deaths is large enough, say $O > 10$, we can use the Normal approximation and obtain a $100(1 - \alpha)$% confidence interval for the SMR:

$$\text{SMR} \pm z_{\alpha/2} \times 100 \times \frac{\sqrt{O}}{E}$$

3.5 Descriptive Measures

3.5.1 Population Pyramid

The age distribution of a population can be shown as a histogram, but since the distribution for males and females is different, a population pyramid is often used, with the two distributions side by side, the age scale vertically, and frequency scale horizontally. See Figure 1 for an example.
3.5.2 Centile Charts

Centile charts are used to express the standard of some variable that changes with time. Typically they plot the 3rd, 10th, 50th, 90th and 97th percentiles.

3.5.3 Reference Range

The reference range is used to describe what values of a variable normal, healthy people are likely to have. We can draw a sample and use these to estimate a reference range (also called a reference interval or normal range). Typically we use 2.5 and 97.5 percentiles to estimate a 95% reference range. This means that 5% of normal healthy people are outside the ‘normal range’.

If we can assume that the variable is normally distributed, we can use the normal distribution to estimate a 95% reference range by $\bar{x} \pm 2s$.

A confidence interval for these limits can be found if the data are assumed normal. The estimates $\bar{x}$ and $s$ are independent, and we use:

$\bar{X} \sim_{\text{approx}} N \left( \mu, \frac{s^2}{n} \right)$

$s \sim_{\text{approx}} N \left( \sigma, \frac{s^2}{2(n-1)} \right)$

So

$\text{var}(\bar{x} - 2s) = \text{var}(\bar{x} + 2s) \approx s^2 \left( \frac{1}{n} + \frac{2}{n-1} \right)$

If $n$ is large, this is approximately $3s^2/n$.

3.5.4 Demographic Life Tables

Life tables are classically calculated to derive a person’s expectation of life and are much used by insurance companies and actuaries. In this form they are also useful in public health medicine. Rather than charting the progress of a group from birth to death, which is impractical for whole populations, we use a cross-sectional approach; we start with present age-specific mortality rates and calculate what would happen to a cohort of people from birth if these age-specific mortality rates applied unchanged throughout their lives.

Figure 1: Population pyramids for England and Wales for 1901 (left) and 1991 (right).
Table 3: Abridged Life Table 2001-2003, United Kingdom

<table>
<thead>
<tr>
<th>Age</th>
<th>Males</th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q_x$</td>
<td>$l_x$</td>
<td>$d_x$</td>
<td>$e_x$</td>
<td>$q_x$</td>
<td>$l_x$</td>
</tr>
<tr>
<td>0</td>
<td>0.0059</td>
<td>100000</td>
<td>586.5</td>
<td>75.9</td>
<td>0.0048</td>
<td>100000</td>
</tr>
<tr>
<td>5</td>
<td>0.0001</td>
<td>99313</td>
<td>12.9</td>
<td>71.5</td>
<td>0.0001</td>
<td>99432</td>
</tr>
<tr>
<td>10</td>
<td>0.0001</td>
<td>99255</td>
<td>10.7</td>
<td>66.5</td>
<td>0.0001</td>
<td>99380</td>
</tr>
<tr>
<td>15</td>
<td>0.0003</td>
<td>99177</td>
<td>25.8</td>
<td>61.6</td>
<td>0.0002</td>
<td>99324</td>
</tr>
<tr>
<td>20</td>
<td>0.0008</td>
<td>98917</td>
<td>80.9</td>
<td>56.7</td>
<td>0.0003</td>
<td>99206</td>
</tr>
<tr>
<td>25</td>
<td>0.0009</td>
<td>98513</td>
<td>84.4</td>
<td>51.9</td>
<td>0.0003</td>
<td>99059</td>
</tr>
<tr>
<td>30</td>
<td>0.0010</td>
<td>98077</td>
<td>94.2</td>
<td>47.2</td>
<td>0.0004</td>
<td>98885</td>
</tr>
<tr>
<td>35</td>
<td>0.0012</td>
<td>97551</td>
<td>116.1</td>
<td>42.4</td>
<td>0.0006</td>
<td>98634</td>
</tr>
<tr>
<td>40</td>
<td>0.0017</td>
<td>96894</td>
<td>164.3</td>
<td>37.7</td>
<td>0.0010</td>
<td>98263</td>
</tr>
<tr>
<td>45</td>
<td>0.0025</td>
<td>95950</td>
<td>236.6</td>
<td>33.0</td>
<td>0.0016</td>
<td>97663</td>
</tr>
<tr>
<td>50</td>
<td>0.0041</td>
<td>94470</td>
<td>383.7</td>
<td>28.5</td>
<td>0.0027</td>
<td>96677</td>
</tr>
<tr>
<td>55</td>
<td>0.0062</td>
<td>92229</td>
<td>574.4</td>
<td>24.1</td>
<td>0.0040</td>
<td>95170</td>
</tr>
<tr>
<td>60</td>
<td>0.0108</td>
<td>88703</td>
<td>958.7</td>
<td>20.0</td>
<td>0.0066</td>
<td>92866</td>
</tr>
<tr>
<td>65</td>
<td>0.0170</td>
<td>83063</td>
<td>1408.0</td>
<td>16.1</td>
<td>0.0104</td>
<td>91700</td>
</tr>
<tr>
<td>70</td>
<td>0.0279</td>
<td>74683</td>
<td>2087.3</td>
<td>12.6</td>
<td>0.0172</td>
<td>83656</td>
</tr>
<tr>
<td>75</td>
<td>0.0484</td>
<td>62342</td>
<td>3017.0</td>
<td>9.6</td>
<td>0.0312</td>
<td>74767</td>
</tr>
<tr>
<td>80</td>
<td>0.0788</td>
<td>45866</td>
<td>3614.2</td>
<td>7.2</td>
<td>0.0530</td>
<td>61363</td>
</tr>
<tr>
<td>85</td>
<td>0.1294</td>
<td>27999</td>
<td>3624.1</td>
<td>5.1</td>
<td>0.0937</td>
<td>43554</td>
</tr>
<tr>
<td>90</td>
<td>0.1901</td>
<td>12176</td>
<td>2314.4</td>
<td>3.7</td>
<td>0.1556</td>
<td>23483</td>
</tr>
<tr>
<td>95</td>
<td>0.2810</td>
<td>3424</td>
<td>962.1</td>
<td>2.7</td>
<td>0.2451</td>
<td>8182</td>
</tr>
<tr>
<td>100</td>
<td>0.3683</td>
<td>521</td>
<td>192.0</td>
<td>2.0</td>
<td>0.3485</td>
<td>1548</td>
</tr>
</tbody>
</table>

A full life table contains the following information for each year of age $x$:

- $l_x$: number of people alive at the start of year $x$
- $d_x$: number of people dying during year $x$
- $q_x = d_x/l_x$: the probability of dying during year $x$
- $p_x = 1 - q_x$: the probability of surviving year $x$
- $e_x$: expectation of life at year $x$: the average number of years left given that they have reached an age of $x$.

$$
e_x = \frac{1}{2} + \frac{l_x}{2} \sum_{i=x+1}^{\infty} l_i$$

Similar methods are also used in clinical research for comparing mortality of different groups, for example survival of comparable groups of patients with breast cancer being treated by alternative regimens. In this case it is possible to follow up an entire cohort from birth to death; see survival data in Section 6.