

## SECTION 3

# MEASURING INEQUALITY BY HEALTH AND DISEASE CATEGORIES (USING DATA FROM ADMINISTRATIVE SOURCES)

This section looks at how death and illness are recorded and measured by administrative data sources.

Administrative data used to identify health inequalities tend to be of one of two types:

- ◆ Mortality data ([Section 3.1](#))
- ◆ Morbidity data from health service activity statistics, disease registers and similar official sources ([Sections 3.2 and 3.3](#))

Two main types of data on health and disease are used to investigate inequalities:

- ◆ Data from administrative sources, including mortality data and health service activity data ([Section 3](#)).
- ◆ Data from surveys of health, based on samples from the general population or from people with particular conditions and/or receiving healthcare ([Section 4](#)).

### 3.1 Mortality Data

Mortality levels are expressed as incidence rates, calculated by:

$$\frac{\text{[observed number of deaths in the group, for specified period]}}{\text{[number of people in the group]} \times \left[ \begin{array}{l} \text{average number of years of those} \\ \text{people alive for specified period} \end{array} \right]}$$

Using official data at a national, regional or local level, this usually implies dividing the number of deaths in a year by the national or regional population in the middle of that year (assuming that half of the deaths take place either side).

One-off studies that attempt to assess the death rates can be divided into longitudinal studies (where a cohort is followed up over a number of years) and cross-sectional studies.

- ◆ In longitudinal studies, the number of person years of exposure can be calculated directly from the data collected in the study.
- ◆ In unlinked cross-sectional studies, the denominator is usually estimated as the number of people in the middle year of the study period times the number of years covered by the study period.

### ***Mortality rates for different groups***

Different choices of groups and periods lead to a wide variety of mortality figures. Those commonly used are:

- ◆ The perinatal mortality rate.
- ◆ The infant mortality rate.
- ◆ The under 65 mortality rate.
- ◆ The 65 plus mortality rate.

Mortality data are commonly used to derive two other indicators: the standardised mortality ratio and the life expectancy.

### **Standardised mortality ratios (SMRs)**

#### *Comparison between different areas, groups or variations through time*

Usually, one wants to compare the mortality level in one area with another, or of one group with another, or for the same group over time. In most cases, when assessing levels of inequality, one is focussing on comparisons between socio-economic strata.

The problem is that groups will differ in respects other than their socio-economic status. In particular, both age and sex are associated strongly with morbidity and, depending on the context, likely mortality. The usual practise, therefore, is to *standardise* for age and sex before making any comparison between socio-economic strata (Box 12).

Whilst using SMRs to make comparisons between groups or over time is standard practice, there are some caveats:

- ◆ Making and interpreting comparisons over time should be done with care, because the size of groups being compared change over time

**Box 12****Standardisation**

A standard set of mortality rates for age and sex groups are used and then applied to the age and sex composition of each of the groups under study. The common aim is to generate a predicted or expected number of deaths.

There are two main methods: the direct and indirect approach. The more common, indirect standardisation involves taking the national population figures and applying them to the age–sex breakdown in the local area to give an expected number. The observed number is then divided by the expected number to give the standardised value. Direct standardised death rates is the number of deaths that would be expected in the standard population if the age–sex specific death rates in the study population prevailed, divided by the size of the standard population.

The ‘standardised ratio’ is then calculated as the actual number of deaths divided by the expected number of deaths and conventionally multiplied by 100.

and the categories used to define the groups may change their meaning and significance. In particular, large variations between populations can give strange results [137].

- ◆ In the usual comparison of SMRs, the numbers of deaths involved is relatively small. This means that any change in the composition of the groups being compared can have a disproportionately large impact upon the relative SMRs, because those that move between groups may be more (or less) likely to die.
- ◆ Standardisation must be used thoughtfully because it may conceal important variations in the size of death or health inequalities according to age, sex, place of residence or other parameters that are used in the standardisation process.
- ◆ It may sometimes be better to present separate estimates for sub-groups of the population than to pool all differences in one standardised figure. For example, it is common to distinguish data for men and for women in nearly all measurements of health inequalities.
- ◆ Inskip [138] states that “standardised ratios can only be compared to the median of the distribution from which they are derived and not between two distributions”. In other words, a standardised ratio is only standardised by reference to its parent distribution, for example, an SMR derived from Scottish data cannot be compared to an SMR derived from Welsh data.

- ◆ When used at a small area level, one should be aware of how small the numbers of deaths are.

SMRs are therefore not without problems. Moreover, in practice, the calculation of the SMR for each occupational social class at a given point in time involves comparison of the occupations as recorded at death with the occupational distribution at the most recent Census. Because the Census now appears to be better able to classify persons to more specific occupations than does the SMRs, there will be differences between the occupational distribution at the most recent Census and at the time of death (numerator–denominator bias) making SMRs unreliable for comparison over time.

In addition to the technical difficulties, anomalies and ‘peculiarities’ occur between different occupational social class classifications and the data do not always generate a clear gradient. Examples of specific problems include:

- ◆ Coding of social class at death may include important biases [139].
- ◆ Cause-specific mortality rates may be distorted by policy decisions about the relative importance of different disease categories using the International Classification of Diseases [139].
- ◆ Social stereotypes may affect coding of occupations, especially for those who have had a career of part-time or temporary jobs.

### **Box 13**

#### **Examples of ‘peculiar’ results found in analysis of these data sets include:**

- ◆ The SMRs of commercial artists (occupation unit 020.21) is 107 whilst that of industrial designers (occupation unit 020.02) is 54.
- ◆ The SMRs of foremen responsible for product inspection and packing (occupational group 136) is 160 (based on 893 deaths) and that of industrial inspectors (occupation group 137) is only 71 (based on 3514 deaths).

### Data sources

Small area mortality data are available from the ONS or from the Public Health Observatories. For further information contact the ONS helpdesk on the ONS website or [www.apho.org.uk](http://www.apho.org.uk). The ONS published a paper on the options for small populations in 2003. Ward figures will be increasingly more available in the future.

## Life expectancy

### *What does life expectancy mean?*

Life expectancy at birth for an area in a particular period is an estimate of the number of years a new-born baby would survive, were he or she to experience the particular area's age-specific mortality rates for that time period throughout his or her life. The figure is calculated by applying those age-specific mortality rates to the progress of a hypothetical cohort of 100,000 year on year. Because the age-specific mortality rates are area and time specific, it reflects the current and previous conditions that have led people to die at different ages during that period.

It is important to emphasise that, life expectancy at birth is *not* the number of years a baby born in the area at that time could expect to live. This is because death rates of the area are likely to change in the future and because many of those born in the area will live elsewhere for at least some part of their lives.

Life expectancy can also be estimated at any age. It is sometimes denoted by a subscript, thus  $LE_{60}$  is the corresponding calculation for 60 year olds. There is no simple relationship between the life expectancies calculated for different ages. For example, if for a particular area, female life expectancy at birth is 80 years, it does not follow that life expectancy for someone aged 75 years is simply a further five years. This is because survival from a particular age depends on mortality rates beyond that age and the mortality rates for 75–79 years olds will have changed.

### ***Proximity to death***

Proximity to death is a recent addition to the lexicon of measures of health status.

Evidence shows that the bulk of healthcare expenditures for an individual is in the last few months of their life. There have even been suggestions that this should be the basis for resource allocation. However, it is difficult to see how this can be assessed prospectively, or how predictors of 'proximity to death' would outperform age distributions.

## **3.2 Morbidity Data from Health Service Activity Statistics, Disease Registers and Similar Official Sources**

Other than Hospital Episode Statistics (HES), no comprehensive national small area health service activity data are collected in England. There are, however, a number of partial sources, though it may be difficult to negotiate access to data containing small area identifiers.

### **3.2.1 Summary of Data Sources**

In this section, we describe the main sources available with brief notes on practical questions of access, availability and data quality.

These sources include:

- ◆ General Practice/Primary Care ([Section 3.3.1](#))
  - General Practice Research Database.
  - Morbidity Statistics from General Practice (MSGP) publications.
  - Prescription Cost Analysis, Department of Health.
  - NHS Direct: triage database systems.
  - Regional Drug Misuse Databases.
  - Adult Dental Health Survey.
- ◆ Acute Sector ([Section 3.3.2](#))
  - the Korner returns utilising Patient Administrative System (PAS) data from health authorities and trusts, comprising in-patient and outpatient data, for the HES.
- ◆ Community Health Sector ([Section 3.3.3](#))
  - the Korner returns utilising PAS data.
- ◆ Disease Registers ([Section 3.3.4](#)).

### **3.3 General Practice/Primary Care Data**

Previously, there was no central data archiving for general practice similar to that for hospital admissions and episode data. The history of general practice computing in England has not encouraged standardisation. The new GMS contract (nGMS) 2004 has led to the availability of more information from primary care. The nGMS contract established the Quality Management and Analysis System (QMAS) which allows payment to be made to GPs under the Quality and Outcomes Framework (QOF).

#### Local sources

Small groups of practices have developed local patient databases for research, but there is no universal requirement for GPs to provide anonymised data on individual patients or consultations.

There are many examples of local projects successfully approaching general practices and primary care trusts for data on the incidence of specific problems or conditions. There may be no alternative to such local approaches if details on the severity of symptoms or clinical outcomes are required. The most promising conditions are those for which have standard care management protocols such as diabetes and asthma.

### National sources

One officially supported data warehousing scheme has taken over the anonymised data collected by one of the main computer suppliers (VAMP) as part of its contractual arrangements with practices, whose original purpose was to assess drug safety through linking prescriptions to subsequent reports of side effects. This is known as the *General Practice Research Database (GPRD)* and is maintained by the Medicines and Healthcare Products Regulatory Agency for the Department of Health. It holds anonymised patient level data on reasons for consultation, diagnosis, prescribing and referrals. The number of contributing practices fluctuates, but in January 1999 there were 419 participating practices covering approximately three million patients. Their website says that “GPRD can be used to optimise decision-making across the health research spectrum including clinical epidemiology, drug safety, disease management and drug utilisation. It will be licensed to academics, regulators, pharmaceutical organisations and research service providers.” Potential users should submit a request to the Scientific and Ethical Advisory Group: Medicines. More details can be found by contacting the Medicines’ Control Agency.

The one comprehensive national resource on general practice activity covers *prescribing*. The *Prescription Pricing Authority (PPA)* keeps considerable detail on the prescribing behaviour of every English GP. Details include the volume/number of items of each drug prescribed and cost. Registered users have access to health authority level (and Primary Care Organization (PCO)) level data; anonymised GP or practice level data may be available to designated projects. These can be linked to data on the ward distribution of practice populations (as used in the past by health authorities for resource allocation) to generate small area maps of prescribing behaviour.

For several decades, the main national source of data on English General Practice have been the decennial surveys (carried out in 1955/1956, 1970/1971, 1981/1982 and 1991/1992) that have supported the series of *Morbidity Statistics from General Practice (MSGP)* publications. The MSGP4 was undertaken jointly by the ONS, the Royal College of General Practitioners and the Department of Health in 1991–1992. All general practitioners from 60 general practices recorded details of every face-to-face contact with their patients over the course of a year, and information on the socio-economic characteristics of all patients on their list was collected through a short interview questionnaire (Box 14). The total list size of the 60 practices was approximately 400,000. The sample of general practices is not representative in that only practices where all GPs agreed to participate in the study were used, and the practice had to collect certain information and use certain software which is not standard practice.

**Box 14****Main topics covered by MSGP4:**

Age, sex, socio-economic characteristics (region, urban/rural residence, housing tenure, marital status, household composition, social class, economic position, ethnic group), smoking status, number of consultations and referrals by ICD code, preventative healthcare, home visits, distance between patient's home and surgery.

The data are available as standard tables and data sets, customised tables and data sets, or as a package of data and analysis software from the ONS. For further information contact the ONS helpdesk.

The triage database systems used by *NHS Direct* are a potentially rich source of information on population morbidity, though it is not known to what extent the data includes a postcode or small area identifier. At present these data are periodically analysed in order to monitor presenting problems and referrals, but it is unclear to what extent the supporting data are publicly available. Please contact the NHS directly for more details.

*Dental care* is covered by Korner returns ([Section 3.3.2](#)), but these are unsuitable for small area analysis. Two decennial surveys are also available. The Adult Dental Health Survey (last conducted in 1998) involves interviews including a dental examination of approximately 5000 adults across the U.K. Because of the relatively small sample size, it is unlikely to generate reliable sub-regional results. A similar survey of children attending state schools in the U.K. generally has a sample exceeding 15,000. The most recent survey was conducted in 1993.

Other main sources of information on dental health are the claims submitted by dentists for NHS funded treatment. Information is published by HA on numbers and costs of courses of treatment, and more detailed data on types of treatment are extracted from a 5% sample of claims. It may be possible to access anonymised individual level data. The utility of these data will obviously depend on the local balance between NHS and private provision, which varies greatly across the U.K.

The *Regional Drug Users Databases* receive data from a number of sources, including local DAT teams, GPs, hospital trusts and support services. The aim is to keep an anonymised record of each contact between drug users and the support services. Results are collated centrally by the Department of Health and published in a biannual bulletin. The regional databases may be prepared to release anonymised data with ward level identifiers for local mapping exercises.

### 3.4 Acute Sector Data

The richest source of data, can be found in the *Korner* returns from health authorities and trusts. These cover almost all aspects of a trust's performance and include details of population morbidity as well as service activity and staffing levels. Returns are presented at trust or primary care trust level, and sometimes for units within trusts. Lower level data cannot be extracted from the central returns, though some of the contributing institutions may have compiled the material from systems that provide small area identifiers. High level summaries of Korner data are available on the Department of Health website.

The *Patient Administration Systems (PAS)* are one of several sources for the Korner returns. Several software suppliers produce PAS but they all have to comply with national standards for record structures. For in-patients, these include fields on length of stay, specialty and some diagnostic details. For outpatients, they will indicate the date of attendance, specialty and who was seen. Downloads from these systems are collated centrally to produce the familiar HES data. HES holds limited amount of clinical data – main and secondary diagnoses; it also contains age, gender, postcode and some details on referrals and outcomes. The data are available at a high level of aggregation from the ONS website.

Trying to retrieve more detailed clinical information from the acute sector can be very difficult. It will almost always require negotiation with individual trusts and, unless one is dealing with one of the minority of units that have computerised clinical records, it will involve extracting paper records from hospital archives.

### 3.5 Community Health Sector Data

The Community Health Sector is undergoing major re-structuring as community trusts are wound-down and responsibility for the majority of their services passes to PCOs. Prior to this re-structuring, data recording was very uneven. Although there are several PAS intended for community trusts, not all trusts used computerised systems; in the non-computerised trusts, the Korner returns are compiled manually.

Mental health trusts provide PAS returns for incorporation in HES data, but the quality of the data is thought to be patchy. There is likely to be some disruption to data recording where mental health trusts are forced to separate from combined community and mental health trusts. The National Programme for IT (NPFIT) aims to implement modern, integrated IT providing more efficient delivery of information.