# 2 Methodology

## 2.1 Overview

The data analysis presented in this report is based on Australian death certificate data. Information on diabetes as the underlying cause of death has been available in Australia since the early part of this century, but, information on associated causes of death has been available electronically only since 1997, when multiple cause of death coding was first introduced. Diabetes-related mortality can now be more fully explored by extensively analysing death certificate data to identify where diabetes is listed as an associated cause of death as well as the underlying cause of death.

This chapter provides an overview of Australian death certificate data, details the mortality and population classifications, and explains the statistical concepts presented in the report.

## 2.2 Death certificate data

Registration of deaths in Australia is the responsibility of the State and Territory Registrars of Births, Deaths and Marriages. Both demographic and cause of death information are recorded on the death certificate. Information on the cause of death is supplied by the medical practitioner certifying the death or by a coroner. Other information about the deceased is supplied by a relative or other person acquainted with the deceased, or by an official of the institution where the death occurred. The State and Territory Registrars of Births, Deaths and Marriages provide the information to the Australian Bureau of Statistics (ABS) for coding of cause of death and compilation into aggregate statistics. The Australian Institute of Health and Welfare (AIHW) also holds these data without unique identifiers in a national mortality database.

Information on cause of death on the death certificate is divided into two parts. Part 1 contains information on conditions leading directly to death and Part 2 contains information on conditions that contribute to death but do not lead directly to death. The conditions listed in Part 1 follow a causal sequence, beginning with the immediate cause of death (the final condition resulting in death), and any other associated causes. The underlying cause of death (the disease or injury initiating the sequence of events leading to death) is generally selected from conditions listed in Part 1 of the death certificate, however in some cases conditions may be selected from Part 2 of the death certificate based on specific selection rules.

### Multiple causes of death

Before 1997, the underlying cause of death was the only condition coded, although information on all causes of death was listed on the death certificate—this was in line with the World Health Organization recommendations. However, it is often difficult to identify a single underlying cause of death where there are multiple chronic diseases present, because a single disease may not adequately describe the cause of death, and the sequence of events leading to death may also be unclear. When only a single underlying cause of death is

coded, any other information on causes of death listed on the death certificate is not available in electronic form.

For this reason, in 1997 all morbid conditions, diseases and injuries listed on the death certificate were coded, thus enabling identification of multiple causes of death for death statistics. This change in coding practice has enabled the identification of diabetes not only when it is the underlying cause of death, but also when it is an associated or contributory cause (to be referred to as 'associated' cause). Although a flag in the mortality database has been used since 1994 to indicate whether diabetes was listed as an associated cause, no information has been available on other associated causes listed on the death certificates. Multiple cause of death coding provides a more complete picture of diabetes as a cause of death.

To allow for more detailed analysis, deaths data for 1997 and 1998 have been combined, as diabetes deaths are relatively small in number in any one year and multiple cause of death coding has been available only for data since 1997. In this report, diabetes is viewed from two perspectives: as the underlying cause of death and as an associated cause of death. Where diabetes is the underlying cause of death or an associated cause of death the term 'diabetes-related' deaths has been used.

## Data quality

The Australian Bureau of Statistics uses a variety of quality-control measures to ensure mortality data are as reliable as possible. These include contacting the certifying doctor to obtain additional information if necessary, check-coding of cause of death, detailed computer editing of data, and checks on the statistical output at the individual and aggregated levels. Nonetheless, issues relating to reliability and validity of cause of death data still exist. Some of these issues include inaccurate completion of death certificates, inaccuracy of diagnoses, specifying the mode of death rather than the underlying cause, not listing all causes of death, variation in interpreting causal sequences and conditions that may have contributed to death, changing perceptions of the causal role of diseases, and poor identification of Aboriginal and Torres Strait Islander peoples.

In addition, there are some more specific issues in using death certificate data for analysing diabetes mortality. They include:

- Death certificate data do not distinguish between Type 1 and Type 2 diabetes.
- Diabetes has been shown to be under-reported on death certificates. Diabetes as a diagnosis is omitted from one in three death certificates of people known to have diabetes. A systematic under-reporting exists depending on the stated cause of death and on the mode of treatment (Whittall et al. 1990).
- The causal role of diabetes in mortality is often unrecognised. It is often difficult for physicians to decide whether diabetes was the cause of the death process or even if it had a contributory role.
- Subjectivity may exist in coding diabetes or other diseases (such as diseases of the circulatory system) as the underlying cause of death.
- The selection of a single underlying cause of death may be difficult in people with multiple chronic diseases.

In the absence of any national cohort studies on diabetes mortality, the death certificate data remain the most comprehensively collected national data pertaining to mortality in Australia.

### Scope and coverage

Registration of deaths is a legal requirement in Australia and compliance is virtually complete. All deaths that occur in Australia are within the scope of the collection, with the exception of deaths of foreign diplomatic personnel.

The mortality statistics in this publication relate to the year of registration of death. Usually about 5–6% of all deaths which occur in one year are not registered until the following year or later. For national and State and Territory statistics, this effect is minimal as the proportion of deaths not registered in the year of occurrence is fairly constant from year to year. Year of registration has also been used in analysing death data for Indigenous Australians, urban, rural and remote areas of Australia and levels of socioeconomic disadvantage.

## 2.3 Classifications

### Cause of death

The mortality classification is based on the International Classification of Diseases, Ninth Revision (ICD-9) (WHO 1977). The ICD-9 coding system was first used in Australia in 1979. At the three-digit level, the codes range from 001 to 999 and E800–E999. Table 2.1 lists the ICD-9 codes used for the particular causes of death included in this report.

### Urban, rural and remote areas

Urban, rural and remote areas are identified in this report using the Rural, Remote and Metropolitan Areas (RRMA) classification (DPIE & DHSH 1994). The RRMA classification assigns each statistical local area (SLA) to one of seven categories which can be re-grouped into three larger zones or areas: urban (metropolitan), rural and remote. The classification takes into account population numbers and an index of remoteness. The three zones or areas are defined as follows:

- Urban area
  - Capital cities
  - Other metropolitan centres (urban centre population ≥ 100,000)
- Rural area (index of remoteness < 10.5)
  - Large rural centres (urban centre population 25,000–99,000)
  - Small rural centres (urban centre population 10,000–24,999)
  - Other rural areas (urban centre population < 10,000)
- Remote zone (index of remoteness >10.5)
  - Remote centres (urban centre population ≥ 5,000)
  - Other remote areas (urban centre population < 5,000).

SLA boundaries may be redrawn between censuses, and at each census area classifications and their population counts are updated. It is important to note that both the size of the SLAs and the distribution of the population within SLAs vary considerably. For example, within a remote SLA there can be areas that are rural rather than remote, and vice versa.

Mortality data for the three larger areas (urban, rural and remote) are presented in this report, as numbers of deaths are too small for accurate analysis in the seven-category classification. All diabetes-related deaths were assigned to these three areas.

**Table 2.1: ICD-9 classification for selected causes of death**

| Cause of death | ICD-9 code |
|---|---|
| Infectious and parasitic diseases | 001–139 |
| Neoplasms | 140–239 |
| Endocrine, nutritional and metabolic diseases and immunity disorders | 240–279 |
| *Diabetes* | *250* |
| Diseases of the blood and blood-forming organs | 280–289 |
| Mental disorders | 290–319 |
| Diseases of the nervous system and sense organs | 320–389 |
| Diseases of the circulatory system (cardiovascular diseases) | 390–459 |
| *Hypertensive disease* | *401–405* |
| *Ischaemic heart disease (coronary heart disease)* | *410–414* |
| *Heart failure* | *428* |
| *Cerebrovascular disease (stroke)* | *430–438* |
| *Peripheral vascular disease* | *441–444* |
| Diseases of the respiratory system | 460–519 |
| Diseases of the digestive system | 520–579 |
| Diseases of the genito-urinary system | 580–629 |
| *Renal failure* | *584–588* |
| Diseases of the musculoskeletal system and connective tissue | 710–739 |
| Injury and poisoning (external causes) | E800–E999 |
| Other[(a)] | 630–676, 680–709, 740–799 |

(a) Includes diseases of the skin and subcutaneous tissue, congenital anomalies, certain conditions originating in the perinatal period, complications of pregnancy, childbirth and the puerperium, symptoms, signs and ill-defined conditions.

## Aboriginal and Torres Strait Islander peoples

The identification of Aboriginal and Torres Strait Islander peoples (Indigenous Australians) is not accurately recorded on death certificates in all States and Territories, and consequently a reliable national picture of Indigenous mortality cannot be obtained. At present, there is considerable variation in the quality of data for Indigenous deaths from State to State. For the years 1997 and 1998, only mortality data for Western Australia, South Australia, the Northern Territory and the Australian Capital Territory are considered to be of sufficient quality for publication, with registration of Indigenous deaths estimated to be over 90% complete in these States and Territories. Due to the very small number of Indigenous deaths in the Australian Capital Territory, data from Western Australia, South Australia and the Northern Territory form the basis of the Indigenous analysis in this report. For comparability, the non-Indigenous estimates also include data only from Western Australia, South Australia and the Northern Territory.

The Indigenous identifier in the mortality database was missing for less than 0.05% of deaths in Western Australia, South Australia and the Northern Territory. These deaths were excluded from the analysis.

### Socioeconomic inequalities

There are no reliable data from death certificates in Australia on levels of socioeconomic disadvantage. Although information on occupation is recorded, it is not adequate for analyses involving older people who are more likely to be retired and who are prone to chronic diseases such as diabetes. This report takes an alternative approach, using an index classifying people according to the average disadvantage of their statistical local area (SLA) of usual residence. The Index of Relative Socioeconomic Disadvantage (IRSD), developed by the Australian Bureau of Statistics, is constructed using principal components analysis. It is derived from social and economic characteristics of the local area such as a low income, low educational attainment, high levels of public sector housing, high unemployment, and jobs in relatively unskilled occupations.

For the years 1997 and 1998, deceased persons were classified into quintiles of socioeconomic disadvantage according to the IRSD for their SLA of usual residence, with quintile 1 including the least disadvantaged households and quintile 5 the most. SLAs were grouped into quintiles so that each quintile contained approximately 20% of the total Australian population.

It is important to note that the index of socioeconomic disadvantage relates to the average disadvantage of all people living in the area. Thus the resultant mortality inequalities will be smaller than if the population were classified using individual socioeconomic status areas defined at a lower level than SLA (e.g. census districts). In other words, these measures of socioeconomic inequality will generally understate the true inequality in mortality at the individual level in Australia.

SLA of usual residence could not be mapped to an IRSD value for less than 0.4% of deaths. These deaths were excluded from the analysis.

## 2.4 Statistical methodology

### Identifying the underlying and associated causes of death

The AIHW National Mortality Database contains information on all causes of death recorded on the death certificate (up to 14 causes are coded) as well as demographic information. The database comprises two components: data on the underlying cause of death and data on all causes of death (including the underlying cause, although this is not separately identifiable from the multiple cause of death data set).

Obtaining information on diabetes as the underlying cause and an associated cause of death involved a two-stage process:

1. Identifying where diabetes was the underlying cause of death and linking these records to the multiple cause of death data set. This enabled identification of all causes of death listed on the death certificate. When diabetes was the underlying cause of death, diabetes was deleted from the multiple cause of death data set, so that only associated causes were identified.
2. Identifying where diabetes was an associated cause of death (from the multiple cause of death data set) and linking these records to the underlying cause of death data set. Diabetes was deleted from the underlying cause of death data set.

## Statistics

This report generally uses three types of mortality indicators—proportions, age-specific rates and age-standardised rates. Where numbers are large enough, proportions are calculated separately for males and females and all are expressed as a percentage. Mortality data for 1997 and 1998 have been combined in all the analyses.

### Proportion of all deaths which are diabetes-related

Diabetes as a proportion of all deaths has been calculated as the total number of diabetes-related deaths divided by the total number of deaths,

$$\text{i.e.} \quad r_i = \frac{d_i}{n_i}$$

where $r_i$ is the proportion of all deaths which are diabetes-related for population group $i$, $d_i$ is the number of diabetes-related deaths for population group $i$, and $n_i$ is the total number of deaths for population group $i$, where $i$ is, for example, New South Wales, remote areas, Indigenous Australians etc.

### Proportion of deaths with diabetes as the underlying cause for each associated cause

This proportion has been calculated as the number of deaths for a particular associated cause where diabetes is the underlying cause divided by the total number of deaths where diabetes is the underlying cause,

$$\text{i.e.} \quad r_{ji} = \frac{d_{ji}}{n_i}$$

where $r_{ji}$ is the proportion of deaths with diabetes as the underlying cause of death with associated cause $j$ for population group $i$, $d_{ji}$ is the number of deaths with diabetes as the underlying cause for associated cause $j$ for population group $i$, and $n_i$ is the total number of deaths with diabetes as the underlying cause of death for population group $i$, where $i$ is, for example, New South Wales, remote areas, Indigenous Australians etc.

### Proportion of deaths with diabetes as an associated cause for each underlying cause

This proportion has been calculated as the number of deaths for a particular underlying cause where diabetes is an associated cause divided by the total number of deaths where diabetes is an associated cause,

$$\text{i.e.} \quad r_{ki} = \frac{d_{ki}}{n_i}$$

where $r_{ki}$ is the proportion of deaths with diabetes as an associated cause of death with underlying cause $k$ for population group $i$, $d_{ki}$ is the number of deaths with diabetes as an associated cause with underlying cause $k$ for population group $i$, and $n_i$ is the total number of deaths with diabetes as an associated cause of death for population group $i$, where $i$ is, for example, New South Wales, remote areas, Indigenous Australians etc.

**Proportion of deaths with diabetes as an associated cause within each disease group**

This proportion has been calculated as the number of deaths for a particular underlying cause where diabetes is an associated cause divided by the number of deaths with that particular underlying cause,

$$\text{i.e.} \quad r'_{ki} = \frac{d_{ki}}{n_{ki}}$$

where $r'_{ki}$ is the proportion of deaths with underlying cause $k$ where diabetes is an associated cause for population group $i$, $d_{ki}$ is the number of deaths with underlying cause $k$ and diabetes as an associated cause for population group $i$, and $n_{ki}$ is the total number of deaths with the underlying cause of death $k$ for population group $i$, where $i$ is, for example, New South Wales, remote areas, Indigenous Australians etc.

**Age-specific diabetes-related death rate**

For each age group, the age-specific diabetes-related death rate has been calculated as the number of diabetes-related deaths in an age group divided by the total number of deaths for that age group,

$$\text{i.e.} \quad r_{mi} = \frac{d_{mi}}{n_{mi}}$$

where $r_{mi}$ is the age-specific diabetes–related death rate for age group $m$ in population group $i$, $d_{mi}$ is the number of diabetes-related deaths for age group $m$ in population group $i$, and $n_{mi}$ is the total number of deaths for age group $m$ in population group $i$, where $i$ is, for example, New South Wales, remote areas, Indigenous Australians etc.

The same method has been used to calculate the age-specific death rate for diabetes as the underlying cause and diabetes as an associated cause of death.

**Age-standardised diabetes-related death rates**

Age standardisation is a method of adjustment to allow for the effect of variation in the population age structure when comparing death rates for different years or different locations (e.g. between States and Territories). This report has used the 'direct' standardisation method, which applies the age-specific death rates for a particular year to a standard population (Armitage & Berry 1994). This produces an estimate of the death rate which would have prevailed in the standard population if it had experienced the age-specific death rates in the year under study. The standard population used is the total estimated 1991 mid-year Australian population,

$$\text{i.e.} \quad ASR_i = \frac{\sum r_{mi} P_m}{\sum P_m}$$

where $ASR_i$ is the age-standardised diabetes-related death rate for population group $i$, $r_{mi}$ is the age-specific diabetes-related death rate for age group $m$ in population group $i$, where $i$ is, for example, New South Wales, remote areas, Indigenous Australians etc., and $P_m$ is the standard population in age group $m$.

## Significance testing

Many significance tests have been performed throughout this report, particularly when comparing different populations, such as comparisons between Indigenous and non-Indigenous Australians, and urban, rural and remote areas of Australia. As proportions have generally been compared from unequal sample sizes, a pooled estimate and variance are calculated. These are used to calculate a *z*-score which is then compared with the normal distribution (Armitage & Berry 1994),

i.e. for comparison of the proportion of all deaths which are diabetes-related:

$$r_i = \frac{d_i}{n_i} \qquad i = 1,2$$

$$r = \frac{d_1 + d_2}{n_1 + n_2}$$

$$z = \frac{r_1 - r_2}{\sqrt{\left[ r(1-r)\left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right]}}$$

where $r_i$ is the proportion of all deaths which are diabetes-related for population group $i$, $d_i$ is the number of diabetes-related deaths for population group $i$, $n_i$ is the total number of deaths for population group $i$, population group 1 is the first population group which is being compared with population group 2, $r$ is the pooled estimate and $z$ is the calculated $z$-score.