

# Data gaps

The purpose of this section is to bring attention to the indicators where data availability or quality precludes reporting, to identify specific data quality/availability issues and to suggest strategies for improving our ability to report.

In the time since the previous version of the framework was published, little has changed because resources have been devoted to other priorities.

Data gaps are discussed here:

- in broad and generic terms;
- by describing the issues for which data sources have not yet been identified; and
- by describing the limitations of existing data that frustrate more complete reporting.

A number of recommendations follow from the discussions.

Currently, available data allows reporting against most of the suggested framework, with only some issues not presently able to be described.

## Generic issues

There are a number of limitations with data used to describe health issues in rural, regional and remote areas, but perhaps the two most serious relate to Indigenous identification and the lack of information available for remote and very remote areas from surveys.

Accurate identification of Indigenous people in data collections is important because it allows:

- description of health issues for Indigenous and non-Indigenous people separately; and
- disentanglement of Indigenous from remote health issues.

Information from surveys such as the National Health Survey is available for rural and regional areas; however, the number of people surveyed in remote areas is frequently insufficient to draw meaningful conclusions.

## Indigenous identification

Reporting particularly for Indigenous (but also for non-Indigenous) people is substantially affected by the accuracy of identification of Indigenous people in data collections.

In many data sets there is under-identification of Indigenous people, which results in reporting of rates that are lower than is likely to be the case in reality. More seriously, there is a strong likelihood that identification in remote areas is more accurate than in major cities.

The consequence of systematic regional differences in the accuracy of Indigenous identification is that higher rates reported in more remote areas may in fact be an artefact of more accurate identification in those areas. For example, better identification of Indigenous people who are admitted to hospital in remote areas may suggest higher rates of admission than for Indigenous people from major cities. The magnitude of this effect has been measured for hospital morbidity data (ABS & AIHW 1999) but details are unclear for many data sources (including national mortality data).

Additionally, there have been changes over time in the tendency for people to identify as Indigenous, with people more likely to identify now than in the past. Consequently, comparison of Indigenous rates for two time periods may show higher rates in the latter period than would have been the case if the likelihood of identifying as Indigenous had remained constant.

In some cases, it is not possible to report for Indigenous and non-Indigenous people because no information about Indigenous status has been collected or the identification is not considered adequate.

In other cases (for example, in many surveys including the ABS National Health Survey), the number of Indigenous people from whom data was collected in each area is so small that meaningful reporting is not possible, especially in the more remote areas and particularly for rarer conditions.

Not only is poor identification of Indigenous status a constraint for the reporting of regional Indigenous health issues, but it also constrains the reporting of non-Indigenous statistics. At the national level this is not a substantial problem because the proportion who are Indigenous is small. However, in regional and especially remote areas, where Indigenous people are more strongly represented, there is the potential for the number of non-Indigenous people in data collections to be over stated. Against this, if identification of Indigenous people is indeed better in more remote areas (as is likely to be the case), then any error in calculated non-Indigenous rates associated with inaccurate identification of Indigenous status will be moderated. This moderation is largely a consequence of the rarity of Indigenous people in less remote areas and their substantial presence in more remote areas.

## **Surveys**

While several sources of data capture every event or individual (e.g. ABS Census data, national mortality data, national hospital morbidity data), some data sources are samples of the population. A characteristic of samples is that the number of events or individuals recorded is limited, and that consequently it can be difficult to draw firm conclusions from the data. This problem is exacerbated when describing rates for sub-populations, for example people who live in regional and especially remote areas. Because of the expense of collecting data, and to maintain the representativeness of the sample, surveys that are designed to describe national rates will usually have only a small number of responses from remote areas, and consequently it may not be possible to describe rates in those more remote areas (e.g. ABS National Health Survey, National Mental Health Survey, National Survey of Income and Housing Costs, and ABS Survey of Disability, Ageing and Carers).

Additionally, some surveys may sample in the more densely settled parts of remote areas, which may result in people from outlying areas being under-represented, potentially biasing results in the more remote areas. It is currently unclear to what extent, if any, this issue biases the results from surveys such as the National Health Survey in remote areas, and perhaps even in rural/regional areas.

Some surveys (for example, the AusDiab study) employ cluster sampling to provide meaningful and cost-effective national data. The AusDiab study measures a number of biomedical factors (e.g. blood pressure and cholesterol) across Australia. As only a few of the clusters are in non-metropolitan areas, it is not possible to generate meaningful and representative results for regional and remote areas.

CATI surveys are those that are conducted by telephone interview and rely on the respondent having a telephone. Telephone ownership rates are likely to be lower in rural and especially remote areas, particularly among poorer people and Indigenous people. It is possible that sampling in these areas may be biased, with under-representation of people from lower socioeconomic groups and may consequently underestimate rates of poor health outcome in those areas. The size of this effect has not yet been determined.

## **Geographic classification**

There are currently several geographic classifications in existence; RRMA, ARIA and ASGC remoteness structure are in common use. These classifications are not equivalent and it is not possible to make direct comparisons between them. As each new classification is developed, it is adopted by various organisations with different levels of enthusiasm. Indeed, for reporting against some data sets, some classifications may never be appropriate or adopted. As reporting against the range of indicators requires the sourcing of data from a wide range of organisations, the data can frequently not be available and not be organised by the preferred geographic classification. Consequently considerable development work may be required in many cases before all the data are available by a single geographic classification.

The use of a single classification, or several complementary geographic classifications, or the use of some form of geocoding would facilitate reporting from disparate data sources.

Irrespective of the geographic classification used, allocation of a geographic category on the basis of the postcode or statistical local area (SLA) has some deficiencies. One limitation stems from the fact that the smallest geographic area identified in many data sets is the postcode or the SLA. The boundaries for small areas such as these can change from time to time, which can complicate the allocation of broad geographic category. Another limitation is that the boundaries of geographic classifications (e.g. ASGC remoteness) can cut across SLA or postcode boundaries, which means that it is frequently not clear whether a person or event recorded in a data set should be allocated to one remoteness category or another. In these cases they have to be allocated a category on the basis of probability. While it is unlikely that this adds any appreciable degree of systematic bias, it is likely that the ability to distinguish differences in rates between areas is diminished.

Some data collections that originally contained a data field identifying postcode or SLA have lost this when aggregated to the national level. Examples of state data aggregated nationally and losing a geographic identifier in the process include workers' compensation data and cervical screening data.

While it is useful to be able to comment on associations between remoteness and health issues, other aspects of geography have been ignored. Although measures of remoteness (e.g. ARIA and ASGC remoteness structure) are partially affected by the size of the local centre, the effect is small. The size of the local town is of great importance in providing services, employment, education and the opportunity for social interaction, all of which affect health issues. Additionally, there are fundamental differences between coastal and inland areas that are not taken into consideration by the standard geographic classifications; for example, coastal areas can frequently be retirement areas while inland areas are less likely to be. Infrastructure and the natural environment in coastal areas may provide people with a greater opportunity to lead a healthy life. These differences are of fundamental importance and should be considered in future work.

The methods of describing non-metropolitan health outlined in this framework rely on presentation of 'average' findings for broad geographic areas. Such methods allow simple

description of complex issues, however, it is clear that substantial differences exist within areas; some communities exhibit good health while others have poor health (e.g. some have high death rates, others lower death rates). An ability (or the opportunity) to describe the differences between communities within broad areas is important in understanding the fundamental reasons behind this.

## Other issues

- Data may be collected in some locations but not others.  
For example, data about food prices and availability is collected by some states from time to time, but not others. Consequently it is not possible to report differences in food prices and availability from area to area.
- Services may be provided by several different organisations or in several different ways.  
For example, primary medical care is provided by GPs through Medicare or DVA funding and contributions from patients, through state-funded salaried medical officers, some salaried AMS doctors, and perhaps by other means as well, although the services provided by each group may not be equivalent. So as to develop a valid appreciation of a population's access to primary medical care, data from each of these sources needs to be considered. Health services are also provided by state and privately employed allied health workers, state-funded community health centres and services, public health units, and so on. A single repository of data for provision of medical, nursing and allied health services would be, from a reporting perspective, very desirable.
- Frequently it may be difficult to compare rates between areas because of different models of care employed in each of the areas.  
For example, GPs are less likely to charge Medicare or DVA in remote areas; non-metropolitan hospitals are more likely to admit patients; and people from rural and remote areas are more likely to attend hospital accident and emergency (A&E) departments for primary care medical consultations than people from major cities. Comparison between areas in such an environment using disparate data sources may, without a great deal of care, result in invalid comparisons between areas.
- Data may be collected differently in different areas and cannot be aggregated.  
For example, the definition of what constitutes 'child abuse' differs substantially from state to state; consequently it is not possible to provide from national child protection data a regional comparison of the rate of child abuse.
- In a number of cases several data sources (which may not be capable of aggregation) may need to be considered (e.g. data from Medicare, DVA and hospital A&E departments describing primary medical care).
- In many cases the data simply may not have been collected, or a suitable data source may not yet have been identified.  
For example, there is no current national monitoring system that can be used to compare the prices and availability of (healthy) food across the country; consequently comparisons may have to rely on the price of three individual broad commodities (food, petrol and housing).
- In some cases, trends data are not available.  
Where data sources have provided data periodically and regularly, it is possible to describe trends for indicators over time. For a number of issues, data has been collected

on only one occasion, and so it is not possible to describe changes over time (e.g. food prices).

In some situations, even though data have been collected on more than one occasion, a change in either the way the data were collected or aggregated, or a change in the questions asked, has made it impossible to compare across time. In other situations, a change over time in the likelihood of an event (e.g. the likelihood of identifying as Indigenous or the likelihood of a person being correctly identified as Indigenous) can make comparison across time invalid.

Changes in coding from one period to the next can make comparisons over time difficult or impossible. For example, changes to the coding of mortality and hospital morbidity (from ICD9 to ICD10) can frequently result in sudden changes to rates, which can be resolved by adjusting the data (a practical but less than desirable solution). Comparison of the percentage of houses rented, owned and being purchased was adversely affected by a change in coding in the ABS Census between 1991 and 1996; consequently it is only possible to present data for 1996 and 2001.

- While it is possible to describe a number of statistics for each broad geographic area, such statistics do not describe what it is actually like to live and work in a rural or remote community. Case studies would provide more of an understanding of the real problems faced and advantages enjoyed by people who live outside major cities.

## Issues for which it is not possible to report and why (by each dimension of the framework)

- 1.1 All identified issues can be resolved.
- 1.2 The effect of migration on the regional prevalence of disability is unclear. A research project may be able to assess the magnitude of the effect.
- 1.3 Calculation of disability-adjusted life years and disability-adjusted life expectancy requires regional age-specific data on the prevalence of disability. This information currently exists only for Australia as a whole and for Victoria; not for 'rural', 'regional' and 'remote' areas.
- 1.4 While it is possible to describe whether regional death rates are higher or lower, it is not currently possible to describe to what extent these are influenced by any regional differences in the effectiveness of medical or surgical interventions, emergency response to trauma and post-operative care.
- 2.1 There are many gaps in the ability to report for environmental health issues in regional and remote Australia. Sources of data that would allow national regional reporting of water quality; sewerage; food quality and availability; housing quality and function; recreational and cultural facilities; workplace safety and conditions; and pollutants (including agricultural chemicals) have yet to be identified or developed. Details of food availability are periodically assessed by some states, while the ability to report on regional differences in workplace safety could be developed using identified data sources in consultation with the states and territories.
- 2.2 Most identified issues can be described, but it is not currently possible to report on whether work is full-time or part-time, on the seasonality of work, or on the numbers who work on Community Development Employment Programs. Reporting of the numbers who are self-employed and who are employees may be possible with development of data from ABS Rural and Regional Statistics National Centre (RRSNC).
- 2.3 Issues for which data sources have not been identified include social issues and measures of social capital, the availability of social and commercial services, the suitability of housing, the availability of public transport and measures of health literacy. There is no available overall measure of the cost of living in rural and remote areas. An overall summary of business activity may be possible with further development of identified data sources. These issues are particularly important as they describe a substantial part of the social environment in which the health of the population develops.
- 2.4 It is possible to report against most identified issues, however a source of information regarding driving practices has not yet been identified.
- 2.5 Data describing blood cholesterol concentrations is available from the National Health Survey, but relies on people having been tested, frequently in the absence of symptoms. Reporting of blood cholesterol levels from this source may be misleading; results from a random survey may be more valid.

- 3.1 It is unclear where details of the time taken to get victims of trauma to appropriate care is available. It is possible that data from ambulance services may be an appropriate source, but considerable development work is likely to be required in order to generate a national data set. In addition, it is not currently clear where data describing exposure to STI education and its effect on sexual practice can be obtained.
- 3.2 Appropriate care and support after medical or surgical intervention can substantially improve post-operative outcomes and reduce the likelihood of adverse health events. Currently it is not clear where information describing regional differences in the quality of rehabilitation or post-operative care is available.
- 3.3 The source and means of reporting the per capita cost of health services, the cost of screening and the ratio of expenditure to positive health outcomes in each area is currently unclear.
- 3.4 A range of issues under the responsive dimension are currently unable to be described by geographic area. These issues include the cultural appropriateness of services for Indigenous people, the confidentiality of health services, choice of provider, emergency department response and waiting times, the lengths of time people have to wait for appointments with allied health workers or dentists and for pathology and imaging results, the percentage of GPs with closed books and people's general level of satisfaction with available health services. In addition, even though it is possible to describe average bulk billing rates in areas, it is not possible to describe the number of people for whom bulk billing is inaccessible (i.e. the number of people living in communities where bulk billing is not available).
- 3.5 Means of describing reduced access due to discrimination, cost (specifically the additional costs borne by non-metropolitan residents in accessing services), or due to services being periodically closed (e.g. overnight or on weekends) have not been developed. While it is possible to describe the average ratio of health workers to population across broad areas, it is not currently possible to describe this statistic for communities within each broad area (which may be perhaps a better description of access for residents). It is unclear how to describe the level and volume of hospital services available to local residents; this capacity has yet to be developed and appropriate data sources identified. It is currently not possible to describe the number of services provided by community mental health workers and by psychologists.
- 3.6 It is currently not possible to describe survival rates in Intensive Care Units, either for people from each geographic area or for hospitals in each geographic area.
- 3.7 All identified issues can be resolved.
- 3.8 All identified issues can be resolved.
- 3.9 It is not currently possible to report the number of students originally from each area who complete tertiary health courses. While it is possible to describe aspects of retention for general practitioners, it is not yet possible to describe retention of non-medical (e.g. allied) health workers. Additionally it is not possible to clearly describe the demands of on-call work (e.g. the number of weekends each year or nights each week GPs spend on call).

# Data sources used in reporting against the indicators and their constraints

## **ABS National Health Survey**

- Remote areas are poorly represented.
- Reporting for remote and non-remote Indigenous populations is possible using the Indigenous supplement to the NHS survey.
- Health conditions are self-reported and may not be entirely accurate.
- Data in rural and remote areas may be biased because of the tendency to sample in population centres.

## **ABS Mental Health and Wellbeing Survey**

- Remote areas are poorly represented.
- It is not possible to report for Indigenous populations because of the small sample size.
- Conditions are self-reported and may not be entirely accurate.
- Data in rural and remote areas may be biased because of tendency to sample in population centres.
- Time series is not possible.

## **DSRU Child Dental Health Survey**

- Indigenous status is not recorded or is recorded well in only a few states.
- Participation is restricted in some states because of the need for payment from patients.

## **DSRU National Oral Health Survey**

- Data are available only for 1987–88.

## **National Notifiable Diseases Surveillance System**

- Not all cases are notified.

## **NPSU national perinatal database**

- No substantial limitations yet identified.

## **ABS Survey of Disability, Ageing and Carers**

- Remote areas are poorly represented.
- It is not possible to report for Indigenous populations because there is no Indigenous identifier.
- People with a severe disability may move to less remote areas to access services, therefore rates derived from the data may be misleading.
- Although data are representative of the national population, it is unclear whether sampling is truly representative of the population living in each area (especially rural and remote areas).

### **AIHW national mortality data**

- It is not possible to report for Indigenous populations because only an estimated 60% of Indigenous deaths are identified as such. It is probable that this figure is lower in cities and higher outside cities, invalidating comparisons between areas. Indigenous identification is more accurate in some states (e.g. South Australia, Western Australia & Northern Territory) than others, but can change over time.
- Changes in the accuracy of identification of Indigenous people over time invalidates time trend analysis for Indigenous and non-Indigenous people.
- It is probable that older people move to less remote areas so as to access services. Older people who do not require services (and who are therefore healthier or more robust), may be less likely to move.

### **ABS population estimates**

- No substantial limitations yet identified.

### **ABS perinatal deaths data**

- Similar issues to those for national mortality data.

### **ABS births data**

- The quality of the Indigenous identifier is poor.

### **ABS Census**

- Only available every 5 years.

### **DEST university commencements/DEST higher education data holdings**

- There is no guarantee that those commencing will complete studies.
- Although the home address of commencing students is likely to be their parents' home address and therefore where they are 'from', there is no guarantee that this is the case.

### **ABS Survey of Income and Housing Costs**

- Remote areas are not represented.
- It is not possible to report for Indigenous populations because of the small sample size.
- Ability to report time trends is limited, with data collected only in 1996 and 1999–2000.
- Although data are representative of the national population, it is unclear whether sampling is truly representative of the population living in each area (especially rural and remote areas).

### **AIHW National Drug Strategy Household Survey**

- Remote areas are poorly represented.
- It is not possible to report for Indigenous populations because of the small sample size.
- Ability to report time trends is currently limited.

#### **Informed Sources P/L petrol prices**

- No substantial limitations yet identified.

#### **ABS indexes of relative retail prices of food**

- The survey was discontinued in 1990.

#### **ABS ATO Australian Business Register**

- Data are not yet available by ASGC Remoteness, RRMA or ARIA.

#### **ABS ATO business income data**

- No substantial limitations yet identified.

#### **ABS National Nutrition Survey**

- The sample is relatively small, with poor coverage in remote areas.
- It is not possible to report for Indigenous populations because of the small sample size.

#### **Australian Study of Health and Relationships (La Trobe)**

- The survey has been conducted only once and may not be repeated.

#### **Australian Childhood Immunisation Register**

- It is not possible to report for Indigenous children because the Indigenous indicator is unreliable.

#### **AIHW labour force data sets**

- There is some reliance that survey participants will list all the places where they work. It is likely that a proportion list only the main location of their work; consequently labour force in more remote areas may be under-enumerated.
- There is no information about the time spent travelling on the job rather than working with patients. Greater need for work-related travel in non-metropolitan areas may affect comparability between areas.
- There is no data for professions that do not require registration.
- It is not possible to describe details of weekends and nights spent on call or working on call.

#### **Medicare/DVA data**

- Medicare/DVA consultations represent a large proportion of provided medical services, but not all of them.

#### **AIHW National Hospital Morbidity Database**

- The rate of hospital admission can be affected by both the need for care and also the hospital's admission policies (which is likely to be affected by remoteness). Consequently it is unclear whether higher rates of admission reflect higher rates of need or greater likelihood of admission.
- Indigenous people are under-identified, but particularly in metropolitan areas. Consequently comparison between areas of admission rates for Indigenous people is likely to be invalid.

#### **HACC minimum data set**

- No substantial limitations yet identified.

#### **ACCMIS warehouse files**

- No substantial limitations yet identified.

**BEACH data**

- BEACH data is poorly represented in remote areas.
- Reporting of reasons for encounter is less reliable than for problem managed.

**AIHW Elective Surgery Waiting Times Data Collection**

- It is not possible to report the waiting times for people who live in each area, although it is possible to report the waiting times experienced by people at hospitals in each area.

**AIHW National Hospital Establishments Database**

- Counts of outpatient occasions of service relate to the area of the hospital (not the home address of the patients), and it is not possible to describe the characteristics (i.e. age, sex and Indigenous status) of the patients.

**Pharmaceutical Benefits Scheme (PBS) data**

- Data are not available to describe the overall rate of prescription. Only details of pharmaceuticals that are collected for concession holders or that are expensive are available. Pharmaceuticals can also be obtained from other sources (e.g. state and territory health departments), however data from these other sources is not available to the same degree of detail.
- Time trend is not possible.

**CSDA minimum data set**

- Data are available only for the location from which the service was provided, not the home address of the client.

## Strategies to overcome data constraints

This framework is a living document. There is an expectation that the framework and reporting can be improved with further development of data sources and statistical capacity, perhaps along the lines of the strategies outlined below.

Consequently, this version of the framework has been updated to reflect issues identified after the first attempt to report against the indicators (AIHW 2005). In this version there are no changes to the structure of the framework, but there are changes in the detail provided for some of the indicators.

This framework should be reviewed and updated by June 2008 to capitalise on experience gained and opportunities identified during the next round of indicator reporting.

### Improvement or validation of existing data sources

1. Consider augmenting the ABS National Health Survey (and other national surveys) so as to give better coverage in remote areas.
2. Assess the magnitude and direction of any bias in the National Health Survey and other similar surveys that may result from a tendency to sample in areas with higher population densities.
3. Assess the amount of bias introduced by sampling in CATI surveys. Are respondents from rural and remote areas more likely to have higher incomes, belong to higher socioeconomic groups and be healthier than non-respondents from these areas?
4. For each region (i.e. by remoteness category), assess the accuracy of identification of Indigenous people in each data collection (particularly the mortality data collection).
5. Further encourage the more accurate identification of Indigenous people in data collections, with the aim of making identification as close to perfect as possible.
6. Assess the extent to which the distribution of people with disabilities is affected by a need for people to migrate to larger centres where they can better access services.
7. Assess the extent to which older people who are in poor health move to less remote areas so as to access services. This report speculates that people who are in poor health, particularly those in older age groups, are likely to move to less remote areas where they can access health services. This goes some way to explaining the lower death rates of older people who live in remote areas, but there is little information available to confirm such a hypothesis.
8. Add questions for Health Labour Force Surveys seeking description of time spent travelling, doing administrative tasks and time spent with patients.

### Development of national data from currently available sources

9. Work with the National Occupational Health and Safety Commission to develop workers' compensation data to provide a national data set capable of reporting for a range of remoteness categories.
10. Develop ABS Rural and Regional Statistics National Centre (RRSNC) data so reporting of income and business data is possible by region (e.g. ASGC Remoteness).

### **Development of the ability to report where data are currently unavailable**

11. There are a number of areas where the ability to report is restricted because of either a lack of identified data sources or the need for considerable development of a meaningful indicator. Focus should be given to the ability to report on:
  - national environmental health issues (e.g. water quality, sewerage, recreational and cultural resources, etc.);
  - regional cost of living;
  - regional availability of healthy food;
  - measures of social capital (e.g. truancy rates, volunteering etc.);
  - services (telephone ownership, banks, access to shops, internet etc.);
  - measures of health literacy;
  - mobility (ease or difficulty of getting to places such as work, services and venues);
  - retrieval times and other details for victims of trauma;
  - quality and effectiveness of follow up care after major medical or surgical intervention;
  - cultural appropriateness of health services for Indigenous people;
  - confidentiality of health data locally;
  - choice of provider;
  - emergency department waiting times;
  - waiting times to consult doctors and allied health workers;
  - waiting times for pathology and imaging results;
  - prevalence of GPs with 'closed books';
  - people's satisfaction with available health services;
  - levels of discrimination against Indigenous people in health services;
  - additional costs borne by residents of non-metropolitan areas in accessing health services;
  - temporarily inaccessible services due to periodic (e.g. weekend or nightly) closure;
  - mental health services provided by community mental health workers and psychologists;
  - survival rates in intensive care units;
  - retention of non-medical health workers (e.g. nurses and allied health workers); and
  - the demands of on-call work for GPs.

### **Capacity building**

12. Develop the ability to report on the effects of coastal versus inland location, town size, and SEIFA on health issues.
13. Develop the capacity to report on the efficiency (including cost efficiency) of health services.
14. Develop geocoding of health and population data as a priority, and also develop a capacity to ascribe aspects of geography (e.g. remoteness, population density, coastal/non-coastal status) on the basis of this specific geographic location. Current organisation of health and population data allows only crude allocation of a remoteness

category or other aspect of geography; this limits the ability to describe aspects of health in rural and remote areas.