Statistical methods for monitoring asthma

The Australian Institute of Health and Welfare is Australia's national health and welfare statistics and information agency. The Institute's mission is *better information and statistics for better health and wellbeing*.

Please note that as with all statistical reports there is the potential for minor revisions of data in this report over its life. Please refer to the online version at <www.aihw.gov.au>.

Statistical methods for monitoring asthma

Australian Centre for Asthma Monitoring

2008

Australian Institute of Health and Welfare

Canberra

Cat. no. ACM 13

A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

ISBN 978 1 74024 746 7

Suggested citation

Australian Centre for Asthma Monitoring 2008. Statistical methods for monitoring asthma. Cat. no. ACM 13. Canberra: AIHW.

Any enquiries about or comments on this publication should be directed to:
Australian Centre for Asthma Monitoring
Woolcock Institute of Medical Research
GPO Box M77
Missenden Road
Camperdown NSW 2050
Phone: (02) 9515 5226 (International +61 2 9515 5226)
Email: acam@asthmamonitoring.org

Published by the Australian Institute of Health and Welfare
Printed by Elect Printing

Contents

Acknowledgments

Rosario Ampon, Patricia Correll, Guy Marks, Leanne Poulos, Anne-Marie Waters and Wei Xuan from the Australian Centre for Asthma Monitoring were responsible for the preparation of this document.

Valuable input was received from members of the Steering Committee for the Australian System for Monitoring Asthma. In particular, we are grateful to David Muscatello, Paul Magnus, Donald Campbell and Kuldeep Bhatia for reviewing the draft.

This publication was funded by the Australian Government Department of Health and Ageing through the National Asthma Management Program.

Abbreviations

ABS	Australian Bureau of Statistics
ACAM	Australian Centre for Asthma Monitoring
ACT	Australian Capital Territory
AGPSCC	Australian General Practice Statistics and Classification Centre
AIHW	Australian Institute of Health and Welfare
ARGPE	Asthma-related general practice encounter
ARIA	Accessibility/Remoteness Index of Australia
ASGC	Australian Standard Geographical Classification
ASMA	Australian System for Monitoring Asthma
ASR	Age-standardised rate
ATC	Anatomical Therapeutic Chemical
BEACH	Bettering the Evaluation And Care of Health
CATI	Computer-Assisted Telephone Interview
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CURF	Confidentialised Unit Record File
DALY	Disability-adjusted life year
DDD	Defined daily dose
DoHA	Australian Government Department of Health and Ageing
ED	Emergency Department
GP	General practitioner
HOIST	Health Outcomes Information Statistical Toolkit
ICD-10-AM	International Classification of Diseases, 10th revision, Australian Modification
ICD-9-CM	International Classification of Diseases, Version 9, Clinical Modification
ICPC-2	International Classification of Primary Care—2nd edition
IRSD	Index of Relative Socioeconomic Disadvantage
MBS	Medical Benefits Scheme
NAEPP	National Asthma Education and Prevention Program
NATSIHS	National Aboriginal and Torres Strait Islander Survey
NHDD	National Health Data Dictionary
NHMD	National Hospital Morbidity Database
NHS	National Health Survey
NSW	New South Wales

NT	Northern Territory
PAF	Population attributable fraction
PBS	Pharmaceutical Benefits Scheme
PIN	Personal identification number
PIP	Practice Incentive Program
Qld	Queensland
RADL	Remote Access Data Library
RD	Rate difference
RPBS	Repatriation Pharmaceutical Benefits Scheme
RR	Rate ratio
SA	South Australia
SAND	Supplementary Analysis of Nominated Data
SE	Standard error
SEIFA	Socio-Economic Index for Areas
SLA	Statistical local area
Tas	Tasmania
Vic	Victoria
WA	Western Australia
WHO	World Health Organization
YLD	Years of life lived with disability
YLL	Years of life lost

Preface

Chronic disease surveillance and monitoring requires the collation, analysis and interpretation of data from a range of administrative, research, technical and specific-purpose data sets. The task is best undertaken within the framework of a set of indicators that have been defined to reflect policy priorities. To be effective, data analysis must take account of the strengths and weaknesses of the available data and, wherever possible, apply internally and externally consistent methods that yield summaries that are readily interpretable by data users.

This statistical procedure manual for analysing disease-specific population data has been prepared in order to document and share the experience of the Australian Centre for Asthma Monitoring (ACAM). Since its formation in 2002, ACAM has undertaken extensive analyses of a wide range of administrative and population health data to monitor asthma in Australia. In this manual, the approaches developed by ACAM for interrogating large population data sets from the point of receipt until delivery of final analyses are documented. It is anticipated that the principles applied in relation to interrogating data for asthma monitoring purposes will also be applicable to other chronic disease areas.

1 Introduction

1.1 Background

This manual of statistical procedures for analysing disease-specific population data has been prepared in order to document and share the experience of the Australian Centre for Asthma Monitoring (ACAM). Since its formation in 2002, ACAM has undertaken extensive analyses of a wide range of administrative and population health data to monitor asthma in Australia. In this manual, the approaches developed by ACAM for interrogating large population data sets from the point of receipt until delivery of final analyses are documented. It is anticipated that the principles applied in relation to interrogating data for asthma monitoring purposes will also be applicable to other chronic diseases.

ACAM is a component of the Australian System for Monitoring Asthma (ASMA) and is a collaborating unit of the Australian Institute of Health and Welfare (AIHW). Its principal role is to monitor asthma using national asthma indicators (listed in Table A.1) using a range of available data sources including:

- National Health Survey
- State Computer-Assisted Telephone Interview (CATI) surveys
- National Mortality Database
- National Hospital Morbidity Database (NHMD) (supplemented by state and territory health department hospital databases)
- Bettering the Evaluation And Care of Health (BEACH) and Supplementary Analysis of Nominated Data (SAND) general practice survey
- Pharmacy Guild Survey
- IMS Health data
- Medicare Australia—Medical Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) data
- Emergency department data collections.

ACAM has developed and/or applied a range of methods and approaches for the analysis of these data. In this document, these methods are described in detail in order to provide a statistical manual that can be used for reference by those involved in analyses of these data. While ACAM's analyses pertain to asthma, it is envisaged that the methods detailed in this manual will be useful in other areas, particularly in the field of chronic disease surveillance.

1.2 Structure of this report

This chapter provides background information about ACAM, the data environment in which it operates, and describes the aims of this report. Chapter 2 outlines the general approaches and common measures that ACAM uses in its statistical analyses. Chapters 3 to 11 describe the main data sources analysed by ACAM and the specific statistical methods used to analyse them.

2 Statistical methods

2.1 General approaches

The following sections provide information about the general approaches used for most data sets that are used in asthma monitoring.

2.1.1 Analysis plan

All projects begin with an analysis plan. The analysis plan states the objectives of the analysis, the data requirements, and the analysis that will be undertaken to achieve the stated objectives (see example in Box 1). An analysis plan is particularly important when a new data source is being analysed or when a new type of information is being sought from a previously used data source.

2.1.2 Data request

Once the research questions are formulated, a formal request should be made to the data custodian to obtain permission to access the data. The data selection criteria and a list of the required variables should be included in the letter of request.

2.1.3 Initial data integrity check and data cleaning

Errors within data sets can occur during data entry, with malfunctions of software or hardware, or when the database used does not have a system that filters errors. However, only data custodians can undertake major data cleaning.

Data users should systematically check frequencies, ranges and distributions of each variable in the data set and consider whether these seem plausible. Implausible or unexpected values should be checked. A discussion with the data custodian(s) or consulting the data dictionary supplied by the data custodian can sometimes help clarify this.

If the percentage of outliers is negligible, then a decision should be made on whether or not to exclude them from the analysis. The implications of including or excluding any outliers should be considered carefully.

The completeness of each variable is an important characteristic of the data. Too many missing values can render an analysis meaningless. A judgment needs to be made about what level of missing data is acceptable.

For coded variables, a check should be made that the actual values in the data conform to the expected code ranges.

Box 1: Analysis plan prepared for examining the use of asthma medications using Pharmaceutical Benefits Scheme (PBS) data

This study investigates the pattern of use of medications for asthma in Australia including the demographic variation in their use. The observed pattern of use is compared with the use recommended in asthma management guidelines.

The study aims to answer the following questions:

1. *How does the use of asthma medications vary with demographic factors, namely age, sex, socioeconomic status and remoteness of residence?*
2. *What is the frequency of their use by individuals in Australia?*
3. *How potent are the inhaled corticosteroids prescribed to most people in Australia, and how does this vary with demographic factors?*
4. *How does the pattern of inhaled corticosteroid use relate to the use of other classes of asthma medication?*

Data analysis will involve the following steps:

- *Data will be aggregated by individuals using the anonymous unique patient identifier.*
- *Total number of defined daily doses (DDDs) of each class of medication received by each individual will be calculated.*
- *Postcodes will be used to assign Socio-Economic Index for Areas (SEIFA) and remoteness classifications.*
- *Distribution of DDDs for each class of medication will be described by age group, sex, socioeconomic status and remoteness.*
- *The independent effects of these sociodemographic variables will be quantified by modelling the data (possibly a linear regression, ordinal logistic regression or log-binomial model).*
- *The proportion of people using different potencies of inhaled corticosteroids will be described.*
- *In subgroups of patients in whom data on short-acting beta agonists and oral corticosteroids are available (concessional patients), the relationship between use of inhaled corticosteroids and these medications will be described including:*
  - *the proportion of people taking short-acting beta agonists who also use regular inhaled corticosteroids*
  - *the subgroup of people who, based on short-acting beta agonist supply, should be using inhaled corticosteroids but are not regularly taking this medication.*

As the drugs used in asthma can also be used for other conditions such as chronic obstructive pulmonary disease (COPD), we will increase the specificity of the analysis for asthma by conducting a subsidiary analysis in individuals aged 5 to 34 years.

2.2 Counts

The simplest descriptive variable derived from population data is a count of the number of individuals, cases or events. Estimates of the number of cases or outcomes of disease are important for assessing the absolute size of the problem in a population and for helping governments decide on the resources that need to be allocated to the disease. Incidence and prevalence are two common measures of disease frequency (Box 2). There are various data sources available that provide estimates of counts for incidence and prevalence. For example, estimates of prevalence can be obtained from sample surveys of the population such as the Australian Bureau of Statistics (ABS) National Health Surveys and the state CATI surveys. Counts for incidence can be obtained from disease registers; counts of the number of hospitalisations can be obtained from hospital admissions data such as the National Hospital Morbidity Database (NHMD); and counts of deaths can be obtained from vital statistics collections such as the National Mortality Database. Finally, counts of the population are obtained from the Census.

Box 2: Common measures

Incidence

The number of new cases (of a disease, condition or event) occurring in a population during a given period.

Prevalence

The number or proportion of people with a certain condition in a population at a given time.

2.3 Population-based rates

Rates are used to describe the incidence of an event or the prevalence of a condition in a population or a population subgroup. Incidence rates refer to the number of new cases occurring in a population over a specified time interval divided by the size of the population at the middle of the time interval (Box 2). Prevalence rates refer to the number of people with a specified condition within a population at a given point in time divided by the size of the population at that point in time (Box 2). For rare events, rates are expressed per 100,000 persons. For less rare events or conditions, other bases (for example per 100 persons or percentage) should be used.

ACAM calculates population-based rates using relevant ABS Estimated Resident Population data, which are provided by the AIHW.

2.3.1 Crude rates

A crude rate is simply the total number of cases or events divided by the total population. A crude rate is useful for comparing the relative impact of a health problem between populations regardless of population size. Crude prevalence rates are calculated by dividing the number of people with a condition in a population in a year by the size of that population at the middle of that year. Crude incidence rates are calculated by dividing the number of new events that occurred in a population in a year by the size of that population

in the middle of that year. The mid-year population is an estimate of the average population during the whole year.

For example, the crude prevalence rate for current asthma per 100 persons based on data from the 2004–05 National Health Survey was:

$$\left(\frac{\text{Estimated number of people with current asthma in 2004 – 05}}{\text{Total Australian population as at December 2004}}\right)\times 100$$

= (2,010,212 / 19,681,539) x 100 = 10.2%.

2.3.2 Age- and sex-specific rates

Many conditions have a much greater impact or burden in some age groups than in others. In particular, some diseases have their greatest impact in older people or in children. There may also be important differences between men and women in the burden of disease. An overall crude rate for the whole population may not be a good indicator of the true burden of disease in specific age-sex groups within the population. One way of overcoming this is to express rates separately for specified age groups and for males and females. In this case the relevant cases or events (for the numerator) are those within the specific age–sex group and the relevant population subgroup (for the denominator) is the specified age–sex group. However, in some cases it may be more appropriate to calculate age-specific rates (that is, the number of cases in each specific age group divided by relevant age-specific population) or sex-specific rates (that is, the number of cases for males (or females) divided by relevant male (or female) population), rather than age-sex-specific rates.

2.3.3 Age-standardised rates

When comparing rates of disease between different periods of time, different geographic areas or different population groups it is important to ensure that the observed differences in crude rates are not simply due to differences in the age-sex structure of the populations. This possibility can be overcome by comparing age-sex specific rates but this requires examining several pairwise comparisons. An alternative approach is age-standardisation. This method provides an overall rate for a population, adjusted to a standard population age structure. There are two methods used for age-standardising: direct and indirect. For the purposes of monitoring national asthma indicators, ACAM has mainly used the method of direct age-standardisation.

Direct age-standardisation

Direct age-standardisation is used when the populations under study are large and the age-specific rates are considered to be reliable.

Age-standardised rates (ASR) are calculated using the following formula:

$$\text{ASR} = \sum(r_i P_i) / \sum P_i$$

where r_i is the age-specific rate in age group i of the study population

P_i is the standard population for the ith age group.

ACAM follows the ABS and the AIHW convention of using the age distribution of the population that existed on 30th June in the most recent Census year ending in '1' as the standard population for direct age-standardisation. Therefore, the Australian population as at June 2001 is currently used as the standard population.

Direct age-standardisation converts the crude overall rate estimate for any given year to the rate that would have occurred in that year if the age structure existing in the standard year had existed in that year. Hence, the age-standardised rates currently reported represent the actual rates that would have occurred if the age structure of the 2001 population had existed in other years. Age-standardised rates calculated using different standard populations cannot be directly compared.

The data in Table 2.1 show the age-specific, crude and direct age-standardised death rates for deaths registered for Australian females in 1997 and 2005 for diseases of the respiratory system. The overall crude death rates were very similar in the two years (50.3 per 100,000 population in 1997 and 50.0 in 2005). However, the age-specific rates in 1997 were higher than the corresponding rates in 2005 for all age groups except 10–14 years and 20–29 years. Examination of the age structure of the female population in 2005 reveals that it is slightly older than that in 1997 (46% aged 40 years and over in 2005 compared with 42% in 1997). Hence, the age-standardised rate in 2005 (38.7 per 100,000 population) is lower than that in 1997 (46.0 per 100,000 population).

The standard error (SE) for an age-standardised rate per 100,000 population is calculated as:

$$\text{se(ASR)} = \sqrt{\left(\sum \left[\left((r_i \times P_i^2)/n_i \right) \times 100000 \right] / P^2 \right)}$$

where r_i = age-specific rate per 100,000 for age group i

n_i = population for age group i

P_i = standard population for age group i

$P = \sum P_i$ = total standard population.

When the number of cases is large, the 95% confidence interval for an age-standardised rate can be calculated as:

$$95\%\ \text{CI} = \text{ASR} \pm [1.96 \times \text{se(ASR)}]$$

When the number of cases is small, confidence intervals for direct age-standardised rates can be calculated using a method developed by Dobson et al. (1991).

Indirect age-standardisation

Indirect age-standardisation is used when populations under study are small or where there is some uncertainty about the stability of age-specific rates, for example, when comparing rates between Indigenous and other Australians. The method removes the influence of age structure, but does not provide a measure of prevalence in terms of a rate. Rather, the measure is a ratio of the number of observed cases to the number expected if the age-specific prevalence rates of the standard population are applied to the study population (Anderson & Rosenberg 1998). It is therefore interpreted as an age-adjusted rate ratio. For example, we can use indirect standardisation to compare the death rate due to asthma during 2001–2003 in

Remote/Very Remote areas with that in *Major Cities* using the *Major Cities* population in 2001–2003 as the standard population. To do this we first calculate the age-specific death rates for asthma for the period 2001–2003 for the standard population (i.e. *Major Cities*) as shown in Table 2.2. These age-specific rates are then applied to the *Remote/Very Remote* population in 2001–2003 to calculate the number of asthma deaths that would be expected to occur in each age group over the 3-year period if the age-specific death rates in the *Remote/Very Remote* population were the same as those in the *Major Cities* population. The expected number of deaths in each age group is then summed to give the total number of deaths expected (19.8) in the *Remote/Very Remote* population for the period 2001–2003. However, there were actually 26 deaths due to asthma observed in *Remote/Very Remote* areas during 2001–2003. Therefore, the ratio of observed to expected deaths is 26/19.8 which equals 1.3. This means that, over the period 2001–2003, the adjusted rate of deaths due to asthma in *Remote/Very Remote* areas was 1.3 times as high as (or 30% higher than) that in *Major Cities*.

Table 2.1: An example of direct age-standardisation: Deaths from diseases of the respiratory system[(a)], females, Australia, 1997 and 2005

	Number of deaths		Australian female population		Age-specific rate per 100,000 population		Australian standard population	Expected number of deaths	
Age (years)	**1997**	**2005**	**1997**	**2005**	**1997**	**2005**	**2001**	**1997**	**2005**
0–4	19	14	630,850	614,367	3.0	2.3	1,282,357	39	29
5–9	5	4	642,606	644,571	0.8	0.6	1,351,664	11	8
10–14	1	4	640,419	678,685	0.2	0.6	1,353,177	2	8
15–19	8	3	623,374	672,774	1.3	0.4	1,352,745	17	6
20–24	4	8	669,732	690,799	0.6	1.2	1,302,412	8	15
25–29	10	10	725,782	675,196	1.4	1.5	1,407,081	19	21
30–34	18	7	715,942	760,169	2.5	0.9	1,466,615	37	14
35–39	19	12	742,738	741,694	2.6	1.6	1,492,204	38	24
40–44	31	27	691,698	772,316	4.5	3.5	1,479,257	66	52
45–49	27	28	642,632	735,378	4.2	3.8	1,358,594	57	52
50–54	55	61	537,223	673,196	10.2	9.1	1,300,777	133	118
55–59	117	97	421,185	620,956	27.8	15.6	1,008,799	280	158
60–64	176	184	363,642	468,669	48.4	39.3	822,024	398	323
65–69	356	265	352,263	389,856	101.1	68.0	682,513	690	464
70–74	530	397	328,631	327,840	161.3	121.1	638,380	1,030	773
75–79	731	658	256,462	301,944	285.0	217.9	519,356	1,480	1,132
80–84	852	1,001	179,857	237,955	473.7	420.7	330,050	1,563	1,388
85 and over	1,728	2,333	149,357	211,956	1,157.0	1,100.7	265,235	3,069	2,919
All ages	**4,687**	**5,113**	**9,314,393**	**10,218,321**	**50.3**	**50.0**	**19,413,240**	**8,937**	**7,503**
Age-standardised rate per 100,000 population					**46.0**	**38.7**			

(a) ICD-10 codes J00–J99

Source: AIHW General Record of Incidence of Mortality (GRIM) books AIHW (2007).

Table 2.2: An example of indirect age-standardisation: Deaths due to asthma, Major Cities and Remote/Very Remote areas, Australia, 2001–2003

Age (years)	*Major Cities*			*Remote/Very Remote*	
	Deaths	Population	Rate per 100,000 population	Population	Expected number of deaths
0–14	17	7,570,308	0.2	384,089	0.8
15–19	9	2,687,954	0.3	96,759	0.3
20–24	15	2,889,493	0.5	100,071	0.5
25–29	17	2,974,158	0.6	118,600	0.7
30–34	21	3,134,621	0.7	128,336	0.9
35–39	19	2,995,762	0.6	122,926	0.8
40–44	31	2,983,797	1.1	117,845	1.2
45–49	21	2,721,547	0.8	102,516	0.8
50–54	28	2,578,887	1.1	92,629	1.0
55–59	56	2,107,150	2.7	73,181	2.0
60–64	45	1,604,471	2.8	55,109	1.5
65–69	49	1,325,717	3.7	39,831	1.5
70–74	51	1,217,247	4.2	31,581	1.3
75–79	71	1,038,504	6.9	21,852	1.5
80–84	77	695,038	11.1	13,514	1.5
85 and over	180	553,269	32.5	10,693	3.5
Total	**707**	**39,077,924**	**1.8**	**1,509,532**	**19.8**

2.4 Case-based rates

For the analysis of events or outcomes that are only relevant to people with a specified condition, such as disease management or disease-specific outcomes, rates may be expressed as case-based rates in which the population with the specified condition is the denominator. For example, the denominator may be based on the number of people with asthma as estimated from the most recent ABS National Health Survey.

Comparing population-based rates with case-based rates demonstrates the extent to which variation in population-based rates (for example, in hospitalisations for asthma or deaths from asthma) are attributable to variation in the prevalence of the condition.

It should be noted that case-based rates are subject to greater uncertainty than population-based rates because there is imprecision in the estimation of the denominator, that is, the prevalence of the condition of interest.

In relation to asthma, case-based rates may also vary according to the definition of asthma that was applied in measuring the denominator. Therefore, case-based rates calculated from different studies are not comparable directly if definitions of asthma are different.

2.5 Disability-adjusted life years

Disability-adjusted life years (DALYs) are a set of measures that have been developed to summarise the burden of disease at a population level (AIHW 2006). The DALY is a summary measure of health that provides a common currency by which many diseases, injuries and risk factors can be compared. A DALY represents a lost year of healthy life. It is calculated by adding future years lived with disability (YLD) for incident cases of the health condition to years of life lost due to premature mortality (YLL):

$$DALY = YLD + YLL$$

2.6 Tests for significance and association

2.6.1 Testing for differences in rates between two population groups

A common objective of data analysis is to compare rates between subgroups of the population, for example, to investigate whether the asthma mortality rate differs between males and females. The hypothesis that asthma mortality rates differ between males and females can be tested using a chi-square test. If the p-value associated with the derived chi-square value is less than 0.05, we can accept that asthma mortality rates do differ between males and females. However, this test does not demonstrate the magnitude of the difference.

The relative or absolute magnitude of the effect of sex on asthma mortality rates may be expressed as the ratio of the rates or the difference between the rates, respectively. For example, the rate ratio (RR) comparing age-standardised asthma mortality rates between males (ASR_m) and females (ASR_f) is calculated as:

$$RR = \frac{ASR_m}{ASR_f}$$

where ASR_m and ASR_f are calculated using the formula for direct age-standardised rates in Section 2.2 above.

For example, the age-standardised asthma mortality rates for males and females in 2003 were 1.21 and 1.75 per 100,000 population, respectively (ACAM 2005). Therefore, the rate ratio comparing the male to female death rates is equal to:

$$1.21/1.75 = 0.69$$

This implies that the age-standardised asthma mortality rate for males in 2003 was 0.69 times (or 31% lower than) the rate for females.

The 95% confidence interval (95% CI) for the RR can be calculated as:

$$95\%\ CI = RR^{(1 \pm 1.96/\chi)}$$

where $\chi = (ASR_m - ASR_f)/\sqrt{se(ASR_m)^2 + se(ASR_f)^2}$ (Parkin et al. 1992)

or as:

$$95\%\ CI = e^{(\ln RR \pm 1.96 \times se(\ln RR))}$$

where se(lnRR) = $\sqrt{\frac{se(ASR_m)^2}{ASR_m{}^2} + \frac{se(ASR_f)^2}{ASR_f{}^2}}$ (Rothman & Greenland 1998).

The rate difference (RD) is calculated as:

$$RD = ASR_m - ASR_f$$

Using the example above, the rate difference in the age-standardised male and female asthma mortality rates in 2003 is equal to:

$$1.21 - 1.75 = -0.54$$

This implies that the absolute rate difference between the age-standardised male and female asthma mortality rates in 2003 was 0.54 deaths per 100,000 population.

The 95% confidence interval (95% CI) for the RD can be calculated as:

$$95\%\ CI = RD \pm 1.96 \times se(RD)$$

where se(RD) = $\sqrt{se(ASR_m)^2 + se(ASR_f)^2}$ (Rothman & Greenland 1998).

2.6.2 Testing for trend

In some situations, the rate of a certain condition might be compared across levels of an ordinal (or ranked) categorical variable. For example, the prevalence rate of asthma could be compared across age groups ranging from youngest to oldest. A chi-square test for trend can be used to show whether there is an increasing or decreasing trend in the rate of asthma by age group. This chi-square test for trend is also called the Mantel Haenszel chi-square test (Rothman & Greenland 1998).

2.6.3 Multiple regression

Multiple regression is used to test the association of one or more specified factors of interest (risk factors or predictors) with an outcome after adjusting for the effect of other factors. The parameter estimates for each of the predictor variables included in the regression model provide a measure of the independent association or correlation between each of the predictors and the outcome. That is, they provide an estimate of how much the outcome varies when you alter a predictor variable, while holding all the other predictor variables constant. Relevant predictors used in ACAM's analyses include age, sex, socioeconomic disadvantage (SEIFA quintile), remoteness (ASGC) and Indigenous status. Regression models may also be used to test whether the effect of one predictor on the outcome differs according to another factor. For example, it is possible to test whether the effect of age group on prevalence of asthma differs between men and women.

Ordinary linear, logistic and Poisson regression are from the suite of 'generalised linear models' (McCullagh & Nelder 1999) that can be applied depending on the type of the outcome variable under study.

Most outcomes used in ACAM's analyses are binary; that is they have only two possible values, usually present or absent. The presence or absence of asthma is one example of a binary outcome variable. Logistic regression is the most commonly used multivariate

method for the analysis of these data. Results derived from logistic regression are expressed as odds ratios, with 95% confidence intervals. Strictly speaking, odds ratios should be interpreted as the relative odds of an outcome between people with and without the relevant risk factor or predictor. Odds ratios are often interpreted as an approximation for the relative risk. However, for diseases such as asthma, which are not rare, this assumption may not be valid (Zhang & Yu 2003). Therefore, odds ratios should not generally be stated as representing a ratio of the risk in the study group to that of the comparison group.

It is possible to estimate the relative risk for a binary outcome directly from a multivariate model. This can be achieved using a log-binomial model (Barros & Hirakata 2003) in which the response variable is assumed to be a binomial proportion and the log of the response variable is modelled so that rate ratios (that is risk ratios) can be calculated directly from the parameter coefficients. The pragmatic limitation on the application of this method is that the estimation of the model parameters can be less stable than that of logistic regression. There are instances when the statistical packages cannot actually fit the log-binomial model. In these circumstances it is necessary to revert to the logistic regression model.

Multiple linear regression is used for continuous outcome variables for which the distribution of residual differences between the outcome observations and the modelled values can be assumed to be normally distributed. In most instances, this applies to continuous variables, such as lung function and some quality of life outcomes, which are either normally distributed themselves or can be transformed to a normally distributed variable. The results derived from this regression modelling are expressed as beta coefficients, which represent the change in the outcome variable for each unit change in the predictor variable.

Poisson regression modelling is another generalised linear modelling technique that can be used to model count data and rates where the numerator is a count, such as numbers of hospitalisations (McCullagh & Nelder 1999). Rate ratios can be obtained from the Poisson regression model coefficients and, including age as a predictor variable, provides an alternative means of indirect age-standardisation. This also allows the flexibility of adjusting for other confounding variables such as sex.

2.6.4 Over-dispersion in generalised linear models

One assumption of Poisson regression is that the variance of the model residuals equals their mean value. Similarly, an assumption of models based on the binomial distribution, such as logistic regression, is that the variance of the residuals is directly proportional to the mean. However, these assumptions do not hold when the data are over-dispersed, for example, when the observed variance of the residuals is greater than the mean value for a Poisson model. Over-dispersion can arise for a range of reasons including clustering within the study population and the existence of unmeasured confounding or explanatory factors. It is relatively more important when the sample size is large (McCullagh & Nelder 1999). Over-dispersion results in under-estimation of variance (and, hence, standard errors of the model coefficients) meaning that parameters appear to be estimated more precisely than they actually are.

One method of detecting over-dispersion in the data is to estimate the dispersion parameter. This is calculated as the Pearson chi-square divided by the residual degrees of freedom. This parameter should be approximately equal to 1 if the data are not over-dispersed, while

values greater than 1 may indicate the presence of over-dispersion. Statistical software packages typically allow estimation of the over-dispersion parameter during modelling.

The best approach to dealing with over-dispersion is to fit a model which adequately reflects clustering and all sources of variance. However, this is not always possible. An alternative is to fit models based on distributions in which the variance includes a term to account for over-dispersion. For example, a beta-binomial model can be fitted instead of a binominal model or a negative binomial model can be fitted instead of a Poisson model. Finally, a simple way to correct the variance estimates for the effect of the presence of over-dispersion is to multiply the initial estimate of variance by the over-dispersion parameter, described in the preceding paragraph (McCullagh & Nelder 1999).

2.6.5 Population attributable fraction

Measures of association, such as rate ratios, relative risks, and odds ratios, describe the magnitude or strength of association between a risk factor and specified outcome. However, taken in isolation, they do not inform the investigator about the impact of that risk factor on the outcome in the population. For example, a very strong risk factor may have little impact if it is very rare and a very weak risk factor may have a substantial impact if it is very common in the population. The population attributable fraction (PAF) combines information on the strength of the risk factor and its prevalence in the population. It can be interpreted as the proportion of the specified disease in the population that is attributable to the risk factor under consideration (Walter 1978).

The PAF is defined as the proportion of the disease cases in a population that would be prevented if an exposure were eliminated, assuming the exposure to be causal (Walter 1978). It is calculated as:

$$PAF = \frac{P_e \times (RR - 1)}{P_e \times (RR - 1) + 1}$$

where P_e = proportion of population exposed to the risk

RR = relative risk or rate ratio of exposure.

ACAM has used population attributable fractions to quantify the impact of asthma on quality of life as measured by general health status, days of reduced activity and psychological distress (Ampon et al. 2005). In this analysis, adjusted rate ratios of self-reported current asthma for having each of the quality of life outcomes were estimated using log-linear models with a Poisson distribution. Data on the prevalence of asthma, the risk factor (or 'exposure') for poor quality of life in this case, were derived from the 2001 National Health Survey.

For example, the prevalence of asthma in 2001 was estimated to be 11.2% and the adjusted rate ratio of asthma for having poor self-assessed health was 1.79. Therefore, the adjusted PAF of asthma for poor self-assessed health was equal to:

(0.112*0.79)/(0.112*0.79 + 1) = 0.081 or 8.1%.

This implies that the presence of asthma accounted for 8.1% of people reporting poor health status in 2001, or in other words, that the proportion of people with poor self-assessed health status attributable to asthma in 2001 was 8.1%.

2.7 Survey weights and adjusting for cluster sampling

Most large surveys are conducted using a sample that is not a simple random sample of the target population. Complex sample designs are implemented to ensure that various population subgroups of interest (or strata) are sampled in sufficient numbers to allow adequate precision of estimates for those subgroups. Furthermore, when the survey is implemented, not all members of the target sample actually respond. Usually response rates differ between various subgroups of the population, making the final sample unrepresentative of the target population. In order to draw inferences about the population as a whole, it is necessary to apply weights to the sample data that reflect the sample design and the pattern of response (or non-response) within population strata.

Survey weights for a particular survey are derived to adjust for the sampling design as well as correcting for non-response and lack of coverage in certain subgroups of the population (ABS 2006a). Each observation is assigned a sample weight which reflects the probability of selection of members of that population stratum.

Many surveys also have a two-stage (or multi-stage) sampling procedure. The most common example of this is the household survey. The first stage of sampling is the household. These are randomly selected according to the survey design. This first stage is sometimes referred to as the cluster (since it represents a cluster of individuals who are eligible for the second stage of sampling). The second stage of sampling is conducted among the individuals within the household. Hence, every individual selection is the result of two consecutive sampling procedures. As individuals within a cluster (for example, a household) are more likely to be alike than two individuals from different clusters, respondents selected in this way can be treated as if they have been simply selected from the population as a whole. This clustered sample design must be taken into account in analysing these data. Surveys such as BEACH and the ABS National Health Surveys have complex survey designs based on stratified cluster sampling. In the case of the BEACH survey, the clusters are general practices. In the case of the National Health Survey, the clusters are households.

ACAM has used the SURVEYMEANS® Procedure in SAS software version 9.1 (SAS Institute Inc. © 1999-2001) to incorporate data on the strata sample weights and the clustered design into the analysis of survey data.

2.8 Mapping to demographic, socioeconomic and geographic classifications

2.8.1 Socio-Economic Indexes for Areas (SEIFA)

Socioeconomic disadvantage is an important determinant of health and disease management (Turrell et al. 2006). Locality of residence is one indicator of an individual's level of socioeconomic disadvantage. The Socio-Economic Indexes for Areas (SEIFA) are four locality-based indices derived from the 2001 Census of Population and Housing to measure

aspects of socioeconomic conditions associated with families and households, personal education qualifications and occupation (ABS 2003a).

ACAM uses the SEIFA Index of Relative Socioeconomic Disadvantage (IRSD) to measure socioeconomic status. The IRSD is derived from data on the social and economic characteristics of the locality. These include low income, low educational attainment, high levels of public sector housing, high unemployment, and jobs in relatively unskilled occupations. Of the four 2001 SEIFA indexes, the IRSD is the most comparable to its 1996 counterpart as it uses the same method and the same variables as the 1996 IRSD (ABS 2004).

The 2001 IRSD is constructed so that relatively disadvantaged areas (for example, areas with many low income earners) have low index values (ABS 2003a). These scores are divided into five quintiles of socioeconomic disadvantage, where 1 = most disadvantaged localities and 5 = most advantaged localities (cut-points for the quintiles used in 2001 are shown in Table 2.3).

Table 2.3: SEIFA quintiles and their corresponding 2001 IRSD score

SEIFA quintile	2001 IRSD score
1st quintile (most disadvantaged)	< 950.53
2nd quintile	950.53–<977.56
3rd quintile	977.56–<1007.04
4th quintile	1007.04–<1059.69
5th quintile (most advantaged)	≥ 1059.69

A score and a quintile are assigned to each postcode. The data file containing index scores and quintiles for each postcode is used to merge with the postcodes in the data set being analysed. In this way, localities of residence, based on postcode, are classified according to quintile of socioeconomic disadvantage.

2.8.2 Remoteness index

Remoteness of residence may influence health and access to health care services. Since 2001, the ABS has included a measurement of remoteness in its Australian Standard Geographical Classification (ASGC). The ASGC remoteness classification is based on the Accessibility/Remoteness Index of Australia (ARIA) which was developed in 1998 by the then Commonwealth Department of Health and Aged Care and the National Key Centre for Social Applications of GIS. It allows a quantitative comparison between 'city' and 'country' Australia (ABS 2001a). Using this classification, postcodes can be categorised into five Remoteness Area Classes: *Major Cities*, *Inner Regional*, *Outer Regional*, *Remote* and *Very Remote*.

To classify each unit record in an analysis data set to a remoteness category, an ABS data file containing postcodes and their corresponding ASGC remoteness category is merged by postcode to each unit record in the analysis data set.

3 National Health Survey

3.1 Background

The National Health Survey (NHS) has been conducted by the ABS across Australia since 1977, roughly every 5 years until 2001 and every 3 years from 2001. Information on health status, use of health services and facilities, and health and lifestyle characteristics is obtained from residents of a sample of private dwellings.

Households from all states and territories are selected using a stratified multi-stage area random sample to ensure that all eligible members of the population within a given state or territory have a probability of selection equal to the state or territory sample fraction. Residents from hospitals, nursing homes, convalescent homes, boarding schools, prisons, single quarters from military establishments, non-Australian military personnel in Australia, overseas visitors, and diplomatic personnel from overseas governments are excluded.

The main strengths of the NHS are that it is a nation-wide sample, with a large number of respondents and a high response rate. In 2004–05, 89% of households selected responded fully (ABS 2006a). However, the information is self-reported, and there are no published data on the accuracy of data collection and responses. Unit record files cannot be released unless they are confidentialised, which results in the loss of small area geographic information.

3.1.1 2004–05 National Health Survey

The 2004–05 NHS was conducted between August 2004 and July 2005 in non-sparsely settled areas of all states and territories of Australia (ABS 2006a). Households (n = 19,501) were sampled according to a specified sampling plan. One adult, aged 18 years or over, and where applicable, one child (aged 0 to 17 years), were surveyed from each selected dwelling, providing a total sample of 25,906 respondents. Parents or guardians were interviewed on behalf of children or, where possible, children aged 15 to 17 years were interviewed in person, with parental consent.

3.1.2 2004–05 National Aboriginal and Torres Strait Islander Health Survey

In addition to the NHS, the ABS also conducted a National Aboriginal and Torres Strait Islander Health Survey (NATSIHS) in 2004–05. The NATSIHS, which was conducted in both remote and non-remote areas of Australia, included a total sample of 10,439 Aboriginal and Torres Strait Islander peoples (ABS 2006b). The response rates for the NATSIHS non-remote and remote samples were 83.4% and 85.5%, respectively. The majority of questions used were the same as those administered in the 2004–05 NHS. However, some asthma-specific questions were not included in the 2004–05 NATSIHS, namely those about respiratory symptoms, type of medication used, nebuliser use or actions taken for asthma (see Table 3.1). Furthermore, information about asthma action plans was only collected in non-remote areas.

Table 3.1: Asthma questions included in the 2001 NHS, 2004–05 NHS and 2004–05 NATSIHS

	2001 NHS		2004–05 NHS	2004–05 NATSIHS	
Asthma questions	General and non-remote Indigenous	Remote Indigenous	General	Non-remote	Remote
Ever diagnosed asthma	✓	✓	✓	✓	✓
Current asthma	✓	✓	✓	✓	✓
Whether has written asthma action plan	✓	×	✓	✓	×
Source of written asthma action plan	×	×	✓	✓	×
Whether has standard asthma action plan	✓	×	✓	✓	×
Whether used pharmaceutical medications for asthma in the last 2 weeks	✓	✓	✓	✓	✓
Number of pharmaceutical medications for asthma in the last 2 weeks	✓	×	✓	×	×
Type of pharmaceutical medication used for asthma in the last 2 weeks	✓	×	✓	×	×
Purpose for using all pharmaceutical medications for asthma in the last 2 weeks	✓	×	×	×	✓
Use of nebuliser to administer asthma medications	✓	×	✓	×	×
Frequency of medications used for asthma in the last 2 weeks	×	×	✓	×	×
Action taken for asthma in the last 2 weeks	✓	✓	✓	×	×
Type of action taken for asthma in the last 2 weeks	✓	✓	✓	×	×

3.2 Access to de-identified unit record National Health Survey data

Since 2001, the ABS has made available two versions of de-identified unit record data from the NHS to be accessed with the approval of the Australian Statistician:

- a basic confidentialised unit record file (CURF) available on CD ROM and on the Remote Access Data Laboratory (RADL)
- an expanded CURF accessible through the RADL.

De-identified unit record data from the 2001 Indigenous Health Survey and the 2004–05 NATSIHS are only available through the expanded CURFs via the RADL.

Only authorised users are able to access these de-identified unit record files. Potentially identifying data (for example locality of birth, age) have been recoded to broad groups (for example region of birth, age group) to ensure that information in the CURF cannot be used to identify an individual. The expanded CURF contains some information that is more detailed than that available in the basic CURF.

The RADL is an online batch database query system, under which microdata are held on a server at the ABS. The RADL allows authorised users to submit programs in SAS, SPSS or Stata software to interrogate, analyse, and model the data, and access the results.

ACAM has approval to access the basic and expanded CURFs for both the 2001 and 2004–05 NHS data, as well as the expanded CURFs for the Indigenous 2001 NHS and the 2004–05 NATSIHS.

A few issues to note when analysing NHS data are:

- Some items within the NHS have changed between surveys. For example, a continuous time series is not available for asthma prevalence due to changes in the method of enumeration of asthma cases.
- Not all questions are asked of all respondents, for example in the 2004–05 NHS self-assessed health status was only asked of people aged 15 years and over.
- For confidentiality reasons, many response categories are recoded to broad categories or not reported at all. For example, only the most commonly used asthma medications were separately identifiable in the 2004–05 NHS unit record data. Further, where a particular medication was not included in the ABS coding scheme it was coded to its generic group.

3.3 Asthma questions included in the National Health Survey

In 2001 and 2004–05, the NHS included a specific set of questions about asthma, although the content differs slightly between the two survey years and between the general and Indigenous surveys (Table 3.1).

3.4 Using National Health Survey data for asthma monitoring

ACAM has used data from the National Health Surveys to:

- monitor relevant national asthma indicators
- describe and monitor patterns and trends in the prevalence of asthma
- describe and monitor patterns and trends in the management of asthma
- examine the impact of asthma on quality of life
- examine the association between asthma status and behavioural risk factors such as smoking, physical activity, and overweight and obesity
- examine comorbidity between asthma and other long-term conditions.

The NHS data fields that were used to monitor the indicators relevant to the prevalence of asthma are described in the following sections.

3.4.1 Prevalence of ever having doctor diagnosed asthma

All participants in the 2001 NHS, the 2001 Indigenous NHS, the 2004–05 NHS and the 2004–05 NATSIHS were asked:

- 'Have you ever been told by a doctor or a nurse that you have asthma?'

Participants who answered 'yes' to this question were classified as 'ever having doctor diagnosed asthma' (Table 3.2).

Table 3.2: Classification of 'ever having doctor diagnosed asthma' based on NHS data, 2001 and 2004–05

		Ever having diagnosed asthma	
Survey/SAS data item	**Categories**	**Yes**	**No**
2001 NHS/ASSTAT (Basic & Expanded CURF)	0 = Not applicable 1 = Ever told has condition, still current and long-term 2 = Ever told has condition, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	ASSTAT = 1, 2 or 3	ASSTAT = 0, 4 or 5
2001 Indigenous NHS/ASSTAT (Expanded CURF)	1 = Ever told has condition, still current and long-term 2 = Ever told has conditions, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	ASSTAT = 1, 2 or 3	ASSTAT = 4 or 5
2004–05 NHS/NHPASTHM (Basic & Expanded CURF)	1 = Ever told has condition, still current and long-term 2 = Ever told has conditions, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	NHPASTHM = 1, 2 or 3	NHPASTHM = 4 or 5
2004–05 NATSIHS/NHPASTHM (Expanded CURF)	0 = Not applicable 1 = Ever told has condition, still current and long-term 2 = Ever told has condition, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	NHPASTHM = 1, 2 or 3	NHPASTHM = 0, 4 or 5

3.4.2 Prevalence of current asthma

NHS participants who reported ever being told by a doctor or a nurse that they have asthma were asked:

- 'Do you still get asthma?'

Participants who answered 'yes' to this question were classified as having 'current asthma' (Table 3.3).

Table 3.3: Classification of 'current asthma' based on NHS data, 2001 and 2004–05

		Current asthma	
Survey/SAS data item	**Categories**	**Yes**	**No**
2001 NHS/ASSTAT (Basic & Expanded CURF)	0 = Not applicable 1 = Ever told has condition, still current and long-term 2 = Ever told has condition, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	ASSTAT = 1 or 2	ASSTAT = 0, 3, 4 or 5
2001 Indigenous NHS/ASSTAT (Expanded CURF)	1 = Ever told has condition, still current and long-term 2 = Ever told has conditions, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	ASSTAT = 1 or 2	ASSTAT = 3, 4 or 5
2004–05 NHS/NHPASTHM (Basic & Expanded CURF)	1 = Ever told has condition, still current and long- term 2 = Ever told has conditions, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	NHPASTHM = 1 or 2	NHPASTHM = 3, 4 or 5
2004–05 NATSIHS/NHPASTHM (Expanded CURF)	0 = Not applicable 1 = Ever told has condition, still current and long-term 2 = Ever told has condition, still current but not long-term 3 = Ever told has condition, not current 4 = Not known if ever told, but condition current and long-term 5 = Never told, not current or long-term	NHPASTHM = 1 or 2	NHPASTHM = 0, 3, 4 or 5

3.5 Statistical methods for NHS data

3.5.1 Calculation of standard errors

Exact standard errors

Exact standard errors and relative standard errors can be calculated for any NHS estimate using the ABS replicate weight method (ABS 2006a). The basic idea behind this method is that representative sub-samples, called replicate groups, are selected repeatedly from the whole sample. The ABS derives replicate weights for each replicate group using the jackknife method of replicate weighting, in which the records in one replicate group are given a weight of zero and the remaining records in the other replicate groups are re-weighted to the survey benchmark population. This process is repeated for each replicate group until each record had as many replicate weights attached to it as there are replicate groups with one of these replicate weights being zero. For example, in the 2004–05 NHS the process was repeated 60 times so that each record had 60 replicate weights attached to it. Using the replicate weight method, the standard error (SE) of an NHS estimate (for example the prevalence of current asthma) is calculated as:

$$SE(y) = \sqrt{(((g-1)/g)\sum_{g}(y(g)-y)^2)}$$

where

g = the number of replicate weights

$y(g)$ = estimate obtained using replicate weights for replicate group g

y = estimate from using full sample weights.

The number of replicate weights attached to NHS records varies according to the survey year and, for the 2004–05 NATSIHS, according to the Indigenous and non-Indigenous sub-samples (Table 3.4).

Table 3.4: Number of replicate weights attached to NHS records

Survey	Number of replicate weights
2001 NHS	30
2004–05 NHS	60
Indigenous 2001 NHS	30
2004–05 NATSIHS—Indigenous sub-sample	250
2004–05 NATSIHS—non-Indigenous sub-sample	60

Indicative standard errors

Because of the large number and diverse nature of estimates which can be produced from the NHS, it is not practicable to calculate separate SEs for each estimate. Therefore, the ABS has published tables of indicative standard errors for the 2001 NHS (both general and

Indigenous) (ABS 2002a; 2003c) and the 2004–05 NHS (ABS 2006a). For the general 2001 NHS and the 2004–05 NHS, these indicative standard errors are available for estimates of various sizes by state as well as for Australia. For the Indigenous 2001 NHS, indicative standard errors are only available for national estimates.

The ABS tables of indicative standard errors can be used to calculate indicative standard errors for estimates of any size between the lowest and highest estimates in the table using the following formula (ABS 2006a):

$$\text{SE of estimate} = \text{lower SE} + \left(\left(\frac{\text{size of estimate} - \text{lower estimate}}{\text{upper estimate} - \text{lower estimate}} \right) \times (\text{upper SE} - \text{lower SE}) \right)$$

3.5.2 Re-scaling NHS weights for significance testing

The participants in the NHS are not a simple random sample of the Australian population. Statistical analysis must take account of the stratified and clustered nature of the sample. This is achieved by the application of weights to each observation or record. The weights that are used in the CURF are constructed in such a manner that the total of the weights for all subjects adds up to the total Australian resident population. This is useful for estimating the number of people in Australia with a given condition or risk factor. However, these weights are not appropriate for assigning the number of degrees of freedom in statistical models or hypothesis tests. For this purpose, the assigned weights must be re-scaled, to reflect the actual sample size used in the survey. The NHS weights can be re-scaled using the following formula (Wilson 1999):

$$\text{Re-scaled NHS weight} = \text{weight for record} \times \frac{\text{unweighted sample size}}{\text{total weighted sample size}}$$

For example, the unweighted sample size in the 2004–05 NHS was 25,906 and the total weighted sample size was 19,681,539. Therefore, the re-scaled weight of any 2004–05 NHS record would be calculated by multiplying the original weight for the record by 25,906/19,681,539. The records from this re-scaled weight will add up to the sample size of the NHS survey (25,906), whereas the original weight will add up to 19,681,539.

4 State health surveys

4.1 Background

Most states and territories have conducted their own health surveys for several years. These are conducted by departments of health and are designed to collect health and related data that are relevant to these departments. In almost all cases, these surveys have adopted a CATI methodology.

Methods of sampling vary with some using random digit dialling and others using the electronic telephone directory (White Pages) or a combination of both. All surveys exclude residents of nursing homes, hospitals, or other institutions. Response rates range from 65% to 90% for adults and around 85% for children, depending on the survey sampling technique and the region surveyed.

CATI surveys are conducted more frequently than the NHS and during 2002 New South Wales, South Australia, Western Australia and Victoria moved to continuous surveys. There is capacity for small area analysis in most surveys, with over sampling in rural areas in many cases.

All information is self-reported. The methods differ slightly from state to state and questions are not standardised, which means that, often, estimates cannot be compared between states. In an effort to address this problem, ACAM undertook an extensive and thorough process to establish reliable questions that could be used consistently in Australian health surveys to monitor national asthma indicators. The recommendations were released in May 2007 in a report entitled *Survey questions for monitoring national asthma indicators* (ACAM 2007b). If these questions are widely adopted, they will enable comparisons and pooling of data from different surveys and, eventually, generate useful data for time trend analysis.

The CATI method of surveying excludes people without a telephone from the target population. These are often people who are poorer, unemployed or living in remote areas. Estimates for the general population are generated by the application of weights to adjust for the differences in probabilities of selection among respondents. In New South Wales, post-stratification weights are also calculated to adjust for the fact that different segments of the population are more likely than others to live in households with a telephone and to take part in a survey, resulting in over- or under-representation of some groups (NSW Health Department 2001).

Whilst most states undertake data accuracy checks on a sample of their records, there is no published data on the accuracy of data collection and response.

4.2 Data available from CATI surveys

State and territory CATI surveys provide a valuable source of information on various asthma indicators relevant to monitoring the prevalence of asthma (Table 4.1) as well as aspects of asthma management, control and quality of life (Table 4.2).

When citing data from CATI surveys it is important to note the differences between states and territories and also between survey years from the same state/territory, where applicable. Slight differences in the way the questions are asked may have an impact on the responses received and, hence, the estimates obtained.

Table 4.1: Asthma definitions used in the state CATI surveys

Ever asthma	Current asthma	Survey(s)
Have you ever been told by a doctor or a nurse that you have asthma?	In the last 12 months, have you had symptoms of asthma? In the last 12 months, have you taken treatment for asthma?	Western Australia Health and Wellbeing Surveillance System
Have you ever had asthma? Was your asthma confirmed by a doctor?	Do you still have asthma?	Social, Environmental and Risk Context Information System South Australian Omnibus Survey WANTS Health and Wellbeing Survey
	Symptoms of asthma include coughing, wheezing, shortness of breath and chest tightness when you don't have a cold or respiratory infection. During the past 12 months, did you have any symptoms of asthma? During the past 12 months, did you take asthma medication that was prescribed or given to you by a doctor? This includes using an inhaler, puffer or nebuliser	South Australian Monitoring and Surveillance System
Have you ever been told by a doctor that you have asthma?	In the last 12 months, have you had asthma symptoms (wheezing, coughing, shortness of breath, chest tightness)?	Victorian Population Health Survey
Have you ever been told by a doctor or at a hospital that you have asthma?	Have you had symptoms of asthma or taken treatment for asthma in the last 12 months?	New South Wales Health Survey (child and adult) Queensland Omnibus Survey (2004)
	Symptoms of asthma include coughing, wheezing, shortness of breath and chest tightness when you don't have a cold or respiratory infection. During the past 12 months, did you have any symptoms of asthma? During the past 12 months, did you take asthma medication that was prescribed or given to you by a doctor? This includes using an inhaler, puffer or nebuliser	Queensland Chronic Disease Survey 2000

Table 4.2: Asthma indicators that can be monitored using CATI data

Indicator	Question(s) used	State/territory
Possession of a written asthma action plan	Do you have written instructions on what to do if asthma is out of control?	SA
	Do you have a written asthma management plan from a doctor on how to treat asthma?	NSW Qld
Smoking in homes where children reside	Do you or the other smokers living in this household – Always or usually smoke outside? Sometimes/usually/always smoke inside?	NSW Child Health Survey
	Which of the following best describes your home situation?	Health and Wellbeing Surveillance System, Western Australia
	My home is smoke free (includes smoking is allowed outside only) People occasionally smoke in the house People frequently smoke in the house	
Quality of life	In general, would you say your health is: excellent, very good, good, fair or poor?	WA Health and Wellbeing Surveillance System Queensland Omnibus Survey NSW Health Survey Victorian Population Health Survey 2003
	Overall, how would you rate your health during the past 4 weeks? Excellent, very good, good, fair, poor or very poor?	NSW Health Survey 2002, 2003
	Physical Component Summary (PCS) Score for SF-12	SA Health and Wellbeing Survey 2000 WANTS Survey 2000 (WA, NT and SA)
	Mental Component Summary (MCS) for SF-12	SA Health and Wellbeing Survey 2000 WANTS Survey 2000 (WA, NT and SA)
	Kessler-10 Psychological Distress Scale	WA Health and Wellbeing Surveillance System, 2004 SA Monitoring and Surveillance system July 2002–June 2004 NSW Health Survey 2003 Victorian Population Health Survey 2003 WANTS Survey WA, NT and SA 2000
	Days in the last 12 months when could not work/study/manage day-to-day activities due to asthma	Qld Chronic Disease Survey
	Number of days lost due to asthma	SA Omnibus
	Days in the last 12 months when asthma interfered with ability to work/study/ manage day-to-day activities	NSW Health Survey
	Did asthma interfere with ability to work/study/manage day-to-day activities last 12 months?	Qld Chronic Disease Survey NSW Health Survey
	Asthma limited the child's usual activities in the last 12 months	NSW Child Health Survey
Asthma control	Number of nights in the last month that sleep has been disturbed by asthma	Qld Chronic Disease Survey NSW Health Survey
	Woken at night	SA Omnibus

Access to these data is usually through published results or aggregated data released to ACAM. However, ACAM has approval to access the New South Wales Health Survey data through the Health Outcomes Information Statistical Toolkit (HOIST). HOIST is essentially a data repository that includes different data sets that can be accessed through the web. It uses SAS as its analytical tool and codes can be submitted remotely. Users can also download the data sets that they require for their analyses. The resulting output or log is automatically displayed on the local PC. Users have to sign a confidentiality agreement with NSW Health to be able to access data through the HOIST website. User's manual and examples of statistical analyses codes can also be downloaded from the website <www.hoist.health.nsw.gov.au>.

5 National Mortality Database

5.1 Background

5.1.1 Registration and enumeration of deaths

In Australia, the registration of deaths is a legal requirement and for this reason the data set is considered nearly complete. The National Mortality Database for each year includes all deaths registered in Australia during that year, including those whose usual place of residence is overseas. Australians who die overseas are not included in this data set (ABS 2002b).

From 1979 until 1 January 1999, deaths in Australia were coded according to the International Classification of Diseases version 9 (ICD-9). Deaths registered from 1 January 1999 were coded using International Classification of Diseases version 10 (ICD-10). Deaths occurring during 1997 and 1998 were coded using both versions of the ICD (that is, dual coded) in order to quantify the impact of the version change on the data.

The second, recent, major change to coding of mortality data in Australia occurred in 1997, with the introduction of automated coding. The use of automated coding allows for the production of multiple causes of death data, improved consistency of coding, and enhanced international comparability in mortality statistics (ABS 2001b).

There are several other factors that need to be considered in interpreting mortality data:

- There is a lag in the processing of death registrations, with approximately 5% of deaths in any calendar year being registered in the following year (ABS 2002b). A variable for 'year of death' is included in the data set and it is preferable to report by year of death in public health settings.
- The identifier for Aboriginal and Torres Strait Islander peoples is not reliable in all states. The ABS advises that for 1990 and later years, data for South Australia, Western Australia and the Northern Territory can be considered sufficiently reliable in identifying Aboriginal and Torres Strait Islander peoples for analysis. Queensland data can be used from 1998 onwards. In addition, the Aboriginal and Torres Strait Islander identifier changed in 1998, from dichotomous (Indigenous/non-Indigenous) to separate identification of Aboriginal and Torres Strait Islander peoples.
- Recorded data on occupation are not useful for examining associations between occupation and cause of death, as only the last occupation is recorded on death certificates and not lifetime employment history.

5.1.2 Deaths due to asthma

The asthma indicators that can be monitored with this data set are:

- age-standardised death rate for asthma among persons aged 5 to 34 years
- age-standardised death rate for asthma among persons of all ages.

Mortality from asthma is a relatively rare event and asthma deaths have halved over the last decade. Changes in mortality attributed to asthma may reflect changes in prevalence, disease severity, treatment, and/or diagnostic fashion. Changes in ICD coding over the years have also been reflected in asthma mortality statistics.

There can be misclassification or overlap between asthma and chronic obstructive pulmonary disease (COPD) at ages where COPD is prevalent and, for this reason, diagnosis of death from asthma is deemed most certain between the ages of 5 to 34 years. Therefore, this age range is commonly used for international comparisons of mortality rates attributable to asthma. However, this age range excludes most asthma deaths, since the majority of these occur in the elderly.

5.1.3 Comparability of death data over time

Comparability of mortality data over time is affected by how data are collected, processed and classified. As noted above (Section 5.1.1), there have been two major changes to the way mortality data are processed and classified. First, since 1997, deaths data have been processed using an automated coding system and second, also from 1997, the classification used to code causes of death was changed to the International Classification of Diseases, revision 10 (ICD-10). An analysis of the dual-coded mortality data for asthma demonstrates that it is one of the conditions most affected by the change from ICD-9 to ICD-10, with an overall comparability factor of 0.75 (ABS 2002b). An analysis by ACAM of dual-coded data (provided by the AIHW Population Health Unit) demonstrated that there is marked age dependence in the comparability factors (Baker et al. 2003). There is virtually no impact of the coding change on the number of deaths assigned to asthma below age 35 years. However, fewer deaths are coded to asthma in the older age groups in ICD-10 than would have been the case using ICD-9.

Table 5.1: Comparability factors for asthma mortality data

Age group	Comparability factor
< 35 years	1.00
35 to 64 years	0.84
65 years and over	0.68

5.2 National Mortality Database data items

When requesting national mortality data from the AIHW, it is necessary to stipulate what years of data are needed as well as the exact disease codes. The data request should also state if the disease codes of interest are required for the underlying cause of death only or for the underlying cause plus multiple causes of death (Box 3). When data are requested over long time periods, it is necessary to include the codes that pertain to the different versions of ICD. The ICD-9 code for asthma is 493 and for ICD-10 the codes for asthma are J45 and J46.

Box 3: Cause of death definitions

Underlying cause of death (UCD)

The disease or injury that initiated the morbid train of events which led to death. This is the cause of death recorded in Part I of the death certificate for which there are no other antecedent causes of death. There must always be an underlying cause. If only one cause is listed in Part I of the death certificate, then this is the underlying cause.

Associated cause of death

Any cause listed on the death certificate other than the underlying cause (ABS 2003b). This includes diseases listed in Part I or Part II of the death certificate.

Multiple causes of death (MCD)

All morbid conditions, diseases and injuries entered on the death certificate (either in Part I or Part II). These include those involved in the morbid train of events leading to death that were classified as either the underlying cause, the immediate cause, or any intervening causes and those conditions that contributed to death, but were not related to the disease or condition causing death (ABS 2003b). For example, 'asthma' as a 'multiple cause' would refer to all instances in which asthma is listed somewhere on the death certificate.

Table 5.2 shows the variables commonly included in National Mortality Database extracts provided to ACAM by the AIHW.

Table 5.2: Variables extracted from the National Mortality Database

Short name	Description	Data categories	Type and length of variable
Administrative data			
Mortality ID	A unique number given to each death record	Mort_ID is comprised of: Year of registration (4 digits) State or Territory of registration number 1 thru 8, (1 digit) District (N.B. dropped in 2001), (2 digits) Registration number 8 digits zero filled to the left	Data type: Numeric Field size: Max. 13
Registration number	The number assigned by the Registrars of Births, Deaths and Marriages at the Registration district level	Layout: NNNNNN	Data type: Numeric Field size: Max. 6
Registration district	A code given to identify each Registrar of Births, Deaths and Marriages	Layout: NN	Data type: Numeric Field size: Max. 2
Certification	A code assigned to identify the professional type of the person who certified the death	Layout: N 1 = Doctor 2 = Coroner/ Government Medical Officer 3 = Other 9 = Not stated	Data type: Numeric Field size: Min. 1
Year	Year in which the death was registered	Layout: yyyy	Data type: Numeric Field size: Max. 4
Month	Month in which the death was registered	Layout: mm	Data type: Numeric Field size: Max. 2
Date of death	Date of death of the person	Layout: yyyymmdd	Data type: Numeric Field size: Max. 8
State	The state/territory in which the death was registered	1 New South Wales 2 Victoria 3 Queensland 4 South Australia 5 Western Australia 6 Tasmania 7 Northern Territory 8 Australian Capital Territory	Data type: Alpha Field size: Min. 2

(continued)

Table 5.2 (continued): Variables extracted from the National Mortality Database

Variable name	Description	Data categories	Type and length of variable
Demographic data			
Sex	The sex of the person	1 Male 2 Female	Data type: Numeric Field size: Max. 1
Date of birth	The date of birth of the person	Yyyymmdd = year/month/day	Data type: Numeric Field size: Min. 8
Age at death	Age at which the person died	Age is in years for persons who were over the age of 1 year. For persons who were under 1 year of age, the age is either in months, days, hours or minutes	Data type: Numeric Field size: Max. 3
Country of birth	The country in which the person was born	**Pre 1991** Layout: NNN **1991 onwards** Layout: NNNN	**Pre 1991** Data type: Numeric Field size: Min. 3 **1991 onwards** Data type: Numeric Field size: Min. 4
Indigenous status	An Aboriginal or Torres Strait Islander is a person of Aboriginal or Torres Strait Islander descent who identifies as an Aboriginal or Torres Strait Islander and is accepted as such by the community in which he or she lives	Layout: N **1980 to 1997** 0 = Not Indigenous, Unknown or Not stated 1 = Indigenous 2 & 3 = Indigenous (NSW 1987 only) 9 = Unknown/Not stated (Vic, WA, NT 1983–1987) **1998 onwards** 0 = Not Indigenous or Unknown 1 = Indigenous 4 = Aboriginal 5 = Torres Strait Islander (TSI) 6 = Both Aboriginal and Torres Strait Islander origin 7 = Not Stated	Data type: Numeric Field size: Min. 1
Area of usual residence (5-digit)	A 5-digit code (State code plus SLA) assigned to the address of the residence/home/residential institution last lived in prior to death	Layout: NNNNN	Data type: Numeric Field size: 5
Area of usual residence (9-digit)	A 9-digit code given to the address of the residence/home/residential institution last lived in prior to death	Layout: NNNNNNNNN	Data type: Numeric Field size: 9

(continued)

Table 5.2 (continued): Variables extracted from the National Mortality Database

Variable name	Description	Data categories	Type and length of variable
RRMA	The Rural, Remote and Metropolitan Areas (RRMA), based on the geographical location of usual residence of the person	Layout: AN, AAAN	Data type: Alpha-numeric Field size: Min. 1
Years resident in Australia	Length of time in years resident in Australia	Layout: yy	Data type: Numeric Field size: Min. 2
Cause of death data			
Underlying cause of death (numeric ICD-7, ICD-8, ICD-9)	(a) The disease or injury which initiated the train of morbid events leading directly to death; or (b) The circumstances of the accident or violence which produced the fatal injury	Layout: NNNN	From 1964–1998: Data type: Numeric Field size: Max. 4
Underlying cause of death (alpha-numeric ICD-10)	(a) The disease or injury which initiated the train of morbid events leading directly to death; or (b) The circumstances of the accident or violence which produced the fatal injury	Layout: (ANNN, ANN)	From 1997 onwards: Data type: Alpha-numeric Field size: Max. 4
Cause of death ICD-10	The diseases or conditions recorded on the death certificate consist of: the cause that led directly to the death (the underlying cause of death); the causes that gave rise to the underlying cause of death; and the causes of death that contributed to the death but were not related to the disease or condition causing it	Layout: ANNN, ANN	From 1997 onwards: Data type: Alpha-numeric Field size: Min. 4
Cause of death number	A number given to the cause of death which relates to its place in the order of causes of deaths listed in the ***Cause of Death Table***	Layout: NN 1 = Underlying cause of death 2 to 20 = Multiple cause of death	Data type: Numeric Field size: Max. 2

Source: The AIHW National Mortality Database Documentation, 2005.

5.2.1 Multiple causes of death data

With multiple causes of death data compiled since 1997, ACAM is able to do an analysis on both the underlying cause and multiple causes of death. The data set includes administrative, demographics and cause of death variables. There were 15 variables for the cause of death, with COD1 as the underlying cause of death. Table 5.3 shows the variables in the data set used to analyse multiple causes of death.

Table 5.3: Variables in the dataset used to analyse multiple causes of death

Variables in the data set	NHDD definition (in the data dictionary)
aborig	An Aboriginal or Torres Strait Islander is a person who identifies as being of Aboriginal or Torres Strait Islander origin
age	Age at which the person died
cob	The country in which the person was born
cod1	The underlying cause of death recorded on the death certificate
cod2–cod15	One of the multiple causes of death recorded on the death certificate
dob	The date of birth of the person
dod	The date of death of the person
mort_id	A unique number given to each death record, can be used to link with the main mortality table
rem_ir	Australian Standard Geographical Classification, based on the geographical location of usual residence of the person, *Inner Regional* Australia
rem_mc	As above, *Major Cities* of Australia
rem_or	As above, *Outer Regional* areas of Australia
rem_r	As above, *Remote* areas of Australia
rem_vr	As above, *Very Remote* Australia
sex	The sex of the person
state	The state or territory in which the death was registered
ures5	A 5-digit code (state code + Statistical Local Area (SLA)) assigned to the address of the residence/home/residential institution last lived in prior to death
ures9	A 9-digit code (state code + stat div + stat subdivision + SLA) given to the address of the residence/home/residential institution last lived in prior to death
year	Year in which the death was registered
year_reside	Length of time (in years) resident in Australia

Descriptive statistics on deaths are calculated according to whether the condition of interest was the underlying cause of death or one of multiple causes of death. The number of associated causes included on records where the condition was the underlying cause of death can also be counted. The proportion of records where the condition was reported as the underlying cause of death relative to the total number where the diagnosis was mentioned as one of multiple causes of death can also be calculated. Age-standardised annual rates of death involving the main condition can be calculated to enable trend analysis, with the main condition either as the underlying or one of multiple causes of death.

Multiple causes of death data can also be used to investigate association among various causes of death. The ratio of the incidence of death involving a number of 'other cause' groups among people who had died with the main condition as one of multiple causes of

death, to the incidence of such deaths in people who had died without the main condition, represents the rate ratio for death involving other diseases associated with the main condition. Standard errors and 95% confidence intervals can also be estimated for each rate ratio.

5.3 Data analysis

All asthma mortality rates, and their corresponding 95% confidence interval, should be age-standardised using the direct method and reported per 100,000 population. Where broad age bands are used (for example, 5 to 34 years) age-standardisation should be conducted within this age band. Rates by subgroup (sex, different age groups, remoteness of area and socioeconomic disadvantage) can be produced based on population rates (with the denominator being the mid-year population estimate for the relevant population subgroup).

There is evidence that the reliability of the attribution of death to asthma is poor in older age groups, where misclassification is common (Jones et al. 1999; Sears et al. 1986; Smyth et al. 1996). There may also be misclassification in very young children. For this reason, a secondary analysis of asthma mortality data is undertaken restricted to the 5 to 34 year age range, where the attribution of the death to asthma is most reliable. There may be similar reasons to restrict the age range in which data are presented for other diseases.

6 National Hospital Morbidity Database

6.1 Background

6.1.1 Enumeration of episodes of hospital care

In all states and territories, administrative and clinical patient data are collected and collated at the hospital level. These data are then aggregated by the state and territory health departments. Each state and territory provides agreed minimum data annually (by financial year) to the AIHW which maintains the National Minimum Data Set for Admitted Patient Care on the National Hospital Morbidity Database (NHMD) (AIHW 2007).

Episodes of hospital care are enumerated as 'separations', which are usually discharges after a period of care in hospital. However, deaths in hospital are also counted as separations, as well as movements within and between hospitals and changes in the clinical intent of care.

The NHMD has patient level data including demographic, administrative and diagnostic information about each hospital separation; however, there are no unique patient identifiers such as names and addresses included in records. Records are included for each separation, not for each patient, so anyone who is admitted and separated more than once in the year (that is, re-admitted) will contribute more than one record. In addition, there are a number of other ways that an individual may record multiple separations. If a person is transferred between hospitals or has a care-type change (with changes in clinical intent, for example from acute to palliative care), this will result in more than one 'separation' being recorded for that individual.

Furthermore, anyone admitted in one year and separated from care in another year will be counted in the year in which they separated. The NHMD includes data relating to people admitted in almost all hospitals (AIHW 2007). Hospitals outside the jurisdiction of a state or territory health authority are excluded (such as hospitals operated by the Department of Defence or Correctional Services).

Whilst National Health Data Dictionary (NHDD) definitions form the basis of the NHMD, there may be variations in admission and separation practices, modes of service delivery and coding practices between the different jurisdictions and from year to year. The AIHW recommends that comparisons between states and territories, reporting years and hospital sectors are made with caution.

Prior to the financial year 1998–99, Indigenous identification data should be analysed with extreme care. From 1998–99 onwards, the information provided for Indigenous status from the Northern Territory, South Australia, Queensland and Western Australia is considered acceptable**.** From 2004–05, data from New South Wales are also considered acceptable.

Due to problems with the reliability of the country of birth codes in the hospital morbidity data, country of birth is only available for 1996–97 and subsequent financial years.

6.1.2 Hospital separations due to asthma

The indicators that can be monitored using hospital separations data are:

- rate of hospital separations for asthma
- rate of hospital patient days for asthma
- rate of hospital re-admissions for asthma within 28 days (not available in the NHMD but could possibly be achieved by probabilistic matching using demographic data).

Hospital separation rate provides information on the incidence of severe exacerbations of asthma. Together with length of stay and total hospital bed-days, these data are also relevant to estimation of the economic burden of asthma.

6.2 Coding of hospital diagnoses

In order to interpret hospital statistics, it is important to understand the process by which hospital diagnoses are coded. Each record includes the principal diagnosis, which is defined as 'the diagnosis established after study to be chiefly responsible for occasioning an episode of admitted patient care, an episode of residential care or an attendance at the health care establishment' (Health Data Standards Committee 2006). This diagnosis is obtained from the hospital medical record following the hospital separation (Box 4) and coded by professional coders at the hospital clinical information (medical records) department.

Box 4: Definitions for hospital data

Hospital separation

A hospital separation is the formal process by which a hospital records the completion of treatment and/or care for an admitted patient. The episode of care may be completed by an admitted patient's discharge, death, transfer to another hospital or change in the type of care.

Patient days

The total number of days for patients who were admitted to hospital for an episode of care and who separated during a specified reference period. A patient who is admitted and separated on the same day is allocated one patient day.

Length of stay

Duration of hospital stay, calculated by subtracting the date the patient is admitted from the day of separation. All leave days, including the day the patient went on leave, are excluded.

Average length of stay

The mean number of days of care for inpatient hospitalisations. Calculated by dividing total patient days in a given period by the total number of hospital separations in that period.

Additional diagnoses may similarly be recorded. These are defined as 'conditions or complaints either coexisting with the principal diagnosis or arising during the episode of admitted patient care, episode of residential care or attendance at a health care establishment' (Health Data Standards Committee 2006). Additional diagnoses provide information on factors which result in increased length of stay, more intensive treatment or

the use of greater resources. However, a documented chronic condition, such as asthma, may not be coded as an additional diagnosis if it is judged not to have required additional patient care during an episode. Conditions that relate to an earlier episode of care and had no bearing on patient management during the current episode should not be included as additional diagnoses. Hence, whether a condition is recorded depends on clinical documentation indicating that the diagnosis meets the criteria for additional diagnosis according to ACS 0002. Changes to the Australian Coding Standards over time may have resulted in variations in the reporting of additional diagnoses (see below).

6.2.1 Comparability of hospital diagnoses over time

Hospital data are currently coded using the International Classification of Diseases version 10 Australian Modification (ICD-10-AM). Prior to this, hospital data were coded using a Clinical Modification of ICD-9 (ICD-9-CM). Since 1 July 1998 New South Wales, the Northern Territory, Victoria and the Australian Capital Territory have used ICD-10-AM, with the remaining states adopting ICD-10-AM on 1 July 1999. A small sample of hospitalisations was subject to a dual coding for the financial year 1995–96. Analysis of these data by ACAM has demonstrated an age trend in the comparability between ICD-9-CM and ICD-10-AM data. As observed in the mortality data, the coding change had little impact on the classification of separations for asthma in people aged less than 35 years. At older ages, ICD-9-CM greatly overestimates hospitalisation rates for asthma compared with ICD-10-AM.

In July 2000, ICD-10-AM second edition was released. With each edition of ICD-10-AM, guidelines were provided to coders. In particular, these guidelines included details regarding the instances in which a condition should be recorded as an additional diagnosis. With the release of ICD-10-AM second edition, the guidelines for coding additional diagnoses were clarified to reduce 'overcoding'. These guidelines emphasised that additional diagnoses should be coded only when they met the criteria for being conditions that were significant in the episode of care. This was implemented to prevent the coding of conditions that were not significant in terms of treatment, diagnostic procedures or increased nursing care and/or monitoring (National Centre for Classification in Health 2005).

In 2005, the National Centre for Classification in Health, which was responsible for the revised guidelines, carried out a time series analysis to assess the impact of the recent revisions on the coding of additional diagnoses (National Centre for Classification in Health 2005). This analysis identified a number of conditions for which the rate of coding as additional diagnoses had substantially declined following the introduction of ICD-10-AM second edition. Generally, these decreases were greatest in the year immediately after the release of the second edition. However, there were further decreases in subsequent years. Among the conditions identified were asthma (ICD-10-AM J45 or J46) and chronic obstructive pulmonary disease (ICD-10-AM J44 only). This study reported that asthma diagnoses fell by 68% between 1999–2000 and 2002–03 and the direction of change was consistent across all states, although the magnitude was greater in New South Wales, Queensland, South Australia and Victoria. In the same period, chronic obstructive pulmonary disease diagnoses fell by 47% consistently across most states, although there was a slight increase in the Northern Territory.

These findings suggest that the recording of additional diagnoses has changed over time since the introduction of revised guidelines with ICD-10-AM second edition from July 2000. The largest change occurred immediately after the release of the second edition, however

further decreases continued in subsequent years. Since there was no dual coding—that is, coding using both the old and new systems—it is not possible to establish with certainty whether the observed changes are the result of real changes in hospitalisation rates or are an artefact introduced by the coding change.

NHMD data are reported by financial year, however, for epidemiological analyses it is often more useful to report by calendar year. The conversion from financial year to calendar year basis will result in the following breaks in ICD coding:

- 1994 to first half of 1998:
  - ICD-9-CM all states and territories
- second half of 1998 to first half of 1999:
  - ICD-9-CM for Queensland, South Australia, Western Australia and Tasmania
  - ICD-10-AM for New South Wales, Victoria, the Northern Territory and the Australian Capital Territory
- second half of 1999 to 2000:
  - ICD-10-AM all states and territories.

6.3 National Hospital Morbidity Database data items

Data from 1993–94 (first data available from the National Hospital Morbidity Database) up to the latest year can be requested from the AIHW. The ICD codes relevant to asthma are: code 493 (ICD-9), and J45 and J46 (ICD-10). Table 6.1 shows the different variables commonly included in National Hospital Morbidity Database extracts provided to ACAM by the AIHW.

Table 6.1: Variables obtained from the National Hospital Morbidity Database

Short name/description	Data categories		Type and length of variable
Establishment data			
Australian state/territory identifier (establishment) State/territory of hospitalisation	1	New South Wales	Type: VARCHAR2 Max size: 3
	2	Victoria	
	3	Queensland	
	4	South Australia	
	5	Western Australia	
	6	Tasmania	
	7	Northern Territory	
	8	Australian Capital Territory	
	9	Other territories (Cocos (Keeling) Islands, Christmas Island and Jervis Bay Territory)	
Establishment—sector A section of the health care industry with which a health care establishment can identify	1	A public hospital, (excluding public psych)	Type: NUMBER Max size: 1
	2	A private hospital	
	4	A public psychiatric hospital	
	5	A private free standing day hospital facility	
	6	A private hospital in TAS (TAS did not separately identify private hospitals and private free-standing day hospital facilities for 2001–02)	
Demographic data			
Sex The sex of the person to whom the episode relates	1	A male person	Type: NUMBER Max size: 1
	2	A female person	
	3	An indeterminate sex category	
	9	Not stated/inadequately described	
Country of birth The country in which the person was born	0–9999 Standard Australian Classification of Countries (SACC, ABS catalogue number 1269.0)		Type: NUMBER Max size: 4
Indigenous status Indigenous status of person according to the following working definition: An Aboriginal or Torres Strait Islander is a person of Aboriginal or Torres Strait Islander descent who identifies as an Aboriginal or Torres Strait Islander and is accepted as such by the community with which he or she is associated	1	Indigenous—Aboriginal but not Torres Strait Islander	Type: NUMBER Max size: 1
	2	Indigenous—Torres Strait Islander but not Aboriginal	
	3	Indigenous—Aboriginal and Torres Strait Islander	
	4	Not Indigenous—not Aboriginal or Torres Strait Islander	
	9	Not stated	
Age The age (in years) of the person at the time of admission	Age is calculated as the difference between birth and admission dates (expressed in years)		Type: NUMBER Max size: 3
Age in days for people aged up to 1 year	Count in number of days up to 1 year. Derived by the AIHW. Calculated as the difference between birth and admission dates		Type: NUMBER Max size: 3

(continued)

Table 6.1 (continued): Variables obtained from the National Hospital Morbidity Database

Short name/description	Data categories		Type and length of variable
The age group of the person based on their age at the time of admission	1	Less than 1 year of age (<1)	Type: NUMBER Max size: 2
	2	Aged 1 to 4 years (1–4)	
	3	Aged 5 to 9 years (5–9)	
	4	Aged 10 to 14 years (10–14)	
	5	Aged 15 to 19 years (15–19)	
	6	Aged 20 to 24 years (20–24)	
	7	Aged 25 to 29 years (25–29)	
	8	Aged 30 to 34 years (30–34)	
	9	Aged 35 to 39 years (35–39)	
	10	Aged 40 to 44 years (40–44)	
	11	Aged 45 to 49 years (45–49)	
	12	Aged 50 to 54 years (50–54)	
	13	Aged 55 to 59 years (55–59)	
	14	Aged 60 to 64 years (60–64)	
	15	Aged 65 to 69 years (65–69)	
	16	Aged 70 to 74 years (70–74)	
	17	Aged 75 to 79 years (75–79)	
	18	Aged 80 to 84 years (80–84)	
	19	Aged 85 or older (85+)	
	20	Age unknown	
Australian state/territory identifier State in which the person usually resides	0	Not Applicable (overseas/other)	Type: NUMBER Max size: 1
	1	New South Wales	
	2	Victoria	
	3	Queensland	
	4	South Australia	
	5	Western Australia	
	6	Tasmania	
	7	Northern Territory	
	8	Australian Capital Territory	
	9	Other territories (Cocos (Keeling) Islands, Christmas Island and Jervis Bay Territory)	
	Null	Not provided	
Statistical Local Area Geographical location of person's usual residence	4-digit Statistical Local Area code (Australian Standard Geographical Classification, Australian Bureau of Statistics catalogue number 1216.0)		Type: NUMBER Max Size: 4

(continued)

Table 6.1 (continued): Variables obtained from the National Hospital Morbidity Database

Description	Data categories	Type and length of variable
Accessibility/Remoteness Index of Australia (ARIA) code Accessibility or remoteness of patient's usual residence. This code is used to score the accessibility or remoteness of the patient's usual residence This variable is available in the NHMD after 1998–99	HA *Highly Accessible*—relatively unrestricted accessibility to a wide range of goods and services and opportunities for social interaction A *Accessible*—some restrictions to accessibility of some goods, services and opportunities for social interaction MA *Moderately Accessible* —significantly restricted accessibility of goods, services and opportunities for social interaction R *Remote* —very restricted accessibility of goods, services and opportunities for social interaction VR *Very Remote* —locationally disadvantaged. Very little accessibility of goods, services and opportunities for social interaction 99 Unknown	Type: VARCHAR2 Max size: 2
Length of stay data		
Year of separation Year that an admitted patient completes an episode of care	This field is hard-coded to the financial year by the AIHW for all separations, e.g. '0102' represents the financial year 2001–02	Type: NUMBER Max size: 4
Separation date Date on which an admitted patient completes an episode of care by either formal separation or by statistical separation	Any date occurring during the collection period	Type: DATE Size: DDMMYYYY
Admission date Date on which an admitted patient commences an episode of care by either formal admission or from statistical admission	Any date prior to, and during, period of data collection	Type: DATE Size: dd/mm/yyyy
Length of stay The length of stay of a patient, excluding leave days, measured in days. Synonymous to 'bed days'	Derived by the AIHW. The length of stay is calculated by the AIHW as the difference between admission and discharge dates less leave days (if any), or 1—depending on which is greater	Type: NUMBER Max size: 5
Same day flag Flag to indicate a patient was discharged on the same date as they were admitted	1 Patient *WAS discharged* on the same date as they were admitted NULL Patient *WAS NOT discharged* on the same date that they were admitted	Type: NUMBER Max size: 1
Clinical and related data		
Principal diagnosis The diagnosis established after study to be chiefly responsible for occasioning an episode of admitted patient care, an episode of residential care or an attendance at the health care establishment, as represented by a code	ICD-10-AM (5th edition) codes	Type: VARCHAR2 Max size: 7
Additional diagnosis 1 to 30 A condition or complaint either co-existing with the principal diagnosis or arising during the episode of care or attendance at a health care establishment	ICD-10-AM (5th edition) codes	Type: VARCHAR2 Max size: 7

(continued)

Table 6.1 (continued): Variables obtained from the National Hospital Morbidity Database

Description	Data categories		Type and length of variable
Procedure code 1 to 31 A clinical intervention that is surgical in nature, and/or carries a procedural risk, and/or carries an anaesthetic risk, and/or requires specialised training, and/or requires special facilities or equipment only available in an acute care setting	Australian Classification of Health Interventions (ACHI) (5th edition)		Type: VARCHAR2 Max size: 8
Procedure block number 1 to 31 Block numbers are assigned to groups of ICD-10-AM procedure codes in order to list codes in a sequential order	ICD-10-AM (5th edition) Block numbers		Type: NUMBER Max size: 4
Admission mode Describes the mechanism by which a person begins an episode of care	1	Admitted patient transferred from another hospital	Type: NUMBER Max size: 1
	2	Statistical admission—episode type change	
	3	Other	
	9	Unknown	
Separation mode Status at separation of person (discharge/transfer/death) and place to which person is released (where applicable)	1	Discharge/transfer to an(other) acute hospital	Type: NUMBER Max size: 1
	2	Discharge/transfer to a Residential Aged Care service, unless this is the usual place of residence	
	3	Discharge/transfer to an(other) psychiatric hospital	
	4	Discharge/transfer to other health care accommodation (includes mother craft hospitals)	
	5	Statistical discharge— type change	
	6	Left against medical advice/discharge at own risk	
	7	Statistical discharge from leave	
	8	Died	
	9	Other (includes discharge to usual residence/own accommodation/welfare institution)	
	0	Unknown/not supplied	
AIHW regrouped Diagnosis Related Groups (DRG) version 5.1 A patient classification scheme which provides a clinically meaningful way of relating the number and types of patients in a hospital to the resources required by the hospital	Australian Refined Diagnosis Related Groups version 5.1		Type: VARCHAR2 Max size: 4
AIHW regrouped Major Diagnostic Category (MDC) version 5.1 Twenty-three mutually exclusive categories into which all possible principal diagnoses fall. The diagnoses in each category correspond to a single body system or aetiology, broadly reflecting the speciality providing care	Australian Refined Diagnosis Related Groups version 5.1		Type: VARCHAR2 Max size: 2

(continued)

Table 6.1 (continued): Variables obtained from the National Hospital Morbidity Database

Description	Data categories		Type and length of variable
Care type A phase of treatment within a hospital stay, indicating the type of care provided Please note that the data categories for care type vary and prior to 2000–01 had different values	1.0	Acute care	
	2.0	Rehabilitation care (that cannot be further categorised)	
	2.1	Rehabilitation care delivered in a designated unit	
	2.2	Rehabilitation care delivered according to a designated program	
	2.3	Rehabilitation care, being principal clinical intent	
	3.0	Palliative care (that cannot be further categorised)	
	3.1	Palliative care delivered in a designated unit	
	3.2	Palliative care delivered according to a designated program	
	3.3	Palliative care, being principal clinical intent	
	4.0	Geriatric evaluation and management	
	5.0	Psychogeriatric care	
	6.0	Maintenance care	
	7.1	Newborn—qualified days only	
	7.2	Newborn—qualified and unqualified days	
	7.3	Newborn—unqualified days only	
	8.0	Other admitted patient care	
	9.0	Organ procurement—posthumous	
	10.0	Hospital boarder	
	11.0	Unknown	

Sources: National Health Data Dictionary Version 13; Data Dictionary: National Hospital Morbidity Database, 2002.

6.3.1 Multiple hospital diagnosis data

The methods used to analyse multiple hospital diagnosis data are the same as those used to analyse multiple causes of death data (see previous section). For analysis of multiple hospital diagnosis data, the principal diagnosis should be treated in the same way as the 'underlying cause of death' in mortality data. Similarly, the additional diagnoses in hospitalisation data should be treated in the same way as the 'associated causes of death' in mortality data. The first diagnosis field reported contains the principal diagnosis followed by the additional diagnoses. Table 6.1 presents the full list of variables in the multiple diagnosis data set.

6.4 Data analysis

Hospital separations data can be investigated across all ages, and for subgroups of the population using similar methods to those detailed for mortality. All rates and the corresponding 95% confidence interval are age-standardised using the direct method and reported per 100,000 population.

In addition, the hospital separation rate per 100 people with asthma can also be calculated. For this purpose, prevalence data for current asthma are derived from the National Health Survey.

Day only admissions ('sday' = 1) are excluded from these analyses.

6.4.1 Seasonal variation

Seasonal variation in hospital separation rates may be an important indicator of environmental influences on health. Further, this variation has important resource implications. Analyses limited to annual counts and rates do not reveal seasonal patterns.

To investigate seasonal variation in hospital admission rates, monthly or weekly deviations from the annual mean may be plotted. The differences between the expected (or average) and the observed number of admissions are the deviations which are plotted to reveal the seasonal pattern. Using this method ACAM has shown that pre-school and primary school-aged children are at increased risk for admission for asthma in February, the first month of the school year (ACAM 2003; ACAM 2005).

6.5 Hospital re-admissions

Data on re-admission to hospital are not currently available in the NHMD. The database contains individual records of hospital separations. In order to identify re-admissions by the same individual, a data field identifying individuals within the data set, that is a linkage key, would be required. There is no such field in the data set.

6.5.1 Analysing hospital re-admissions

The AIHW NHMD does not contain a unique record identifier which would allow tracking of re-admissions. In databases where the admitted patient care data does have a medical record number, which is assigned to each patient at the time of admission to a hospital, that record number may be used to track re-admissions. However, patients who are admitted to one hospital and then re-admitted to another hospital cannot usually be identified this way. Some states, territories or regional health services have systems to uniquely identify patients across different hospital facilities, and there is a national move toward more unique patient identifiers. However, this has not yet been implemented.

Another option for identifying records of the same individual is the use of the 'unplanned re-admission flag' which is currently reported in most data sets by the admission clerk based on self-reported information about past hospital admission in the last 28 days, obtained at the time of admission. This information is not included in the AIHW NHMD. However, the validity of this field is uncertain as it relies on the individual to self-report that an admission is a re-admission.

6.5.2 Data linkage

A further possibility is to use patient identifying information to link records for the same individual. This is possible at the state and territory level where patient identifying

information, such as names and addresses, is available and is currently undertaken routinely in New South Wales and Victoria. However, the only potentially identifying variables included in the NHMD are date of birth and postcode (neither of which are supplied without specific ethics approvals and approval from individual states and territories), and sex. None of these items is definitely identifying for an individual in most cases.

Western Australia is currently the only state to use unique patient identifiers. Each person attending a metropolitan hospital in Western Australia is assigned a unique medical record number, although this does not extend to rural areas (E Lloyd [WA Department of Health], pers. comm., 21 August 2002).

In 2006, ACAM published the results of a study to assess the validity of linkage using various sets of potentially identifying information in hospitalisation data (Ringland et al. 2006). This study used hospital data from New South Wales for which the principal diagnosis was asthma. The study compared linkage methods that used many personally identifying variables and probabilistic techniques ('gold standard' linkage) with linkage that used only exact matches of the limited variables that were included on the NHMD (date of birth, postcode and sex). The study demonstrated that linkage using the variables on the NHMD was able to identify 95% of the same matches that were linked in the gold standard linkage. The sensitivity was probably improved through the restriction to records where the principal diagnosis was asthma. It is, therefore, not clear if these results would be applicable to other conditions.

At the time of this report, this linkage method had not been approved to apply nationally; however, this seems the most feasible method for estimating re-admissions nationally in the absence of state-wide unique patient numbers.

7 Emergency department data

Currently, only three states can provide emergency department (ED) data with diagnostic information: Victoria, New South Wales and Western Australia. The absence of diagnostic information limits the usefulness of ED data collected in other states and territories for disease surveillance.

Victoria collects data from all 24 hour hospital EDs. In New South Wales, data are collected from more than 80 of the 150 EDs in that state. Coverage is better for metropolitan than for rural areas. Western Australia includes data from all EDs in metropolitan Perth only.

There are a number of issues to be considered with using ED data for asthma monitoring. The relationship between presentations to the ED and GP for asthma is unclear. One key difference is severity of disease, with people having an acute, severe attack more likely to present to the ED than their GP. However, this may not necessarily be the case in areas where GP services are limited, costly or where after hours services are not available; in these instances the ED may be used as a GP service. This is particularly relevant for rural areas. Variation in ED attendances might be related to variation in the severity of asthma or variation in the proportion of cases attending EDs. Therefore, these two data sources should ideally be examined in tandem, but separation of the effects one has on the other will be difficult.

Not all hospitals report provisional diagnosis information consistently. The lack of universal coverage and a consistent coding system limits the value of this data source for long-term data monitoring. An advantage of this data in New South Wales is its timeliness. Data can be accessed within one month and can, therefore, be used to monitor short-term changes in attendances, which may reflect the acute effect of environmental changes.

8 Bettering the Evaluation And Care of Health general practice survey

8.1 Background

The Bettering the Evaluation And Care of Health (BEACH) program is based on data collected using a continuous survey of GPs that began in 1998. It is run by the Australian General Practice Statistics and Classification Centre (AGPSCC), formerly the General Practice Statistics and Classification Unit (GPSCU), a collaborating unit of the Family Medicine Research Centre of the University of Sydney and the AIHW. A modified classic synchronised sampling procedure is used to select a random sample of GPs from the Medicare Australia database (Britt et al. 2001). Eligible GPs who agree to participate record information on BEACH data collection forms for 100 consecutive patient encounters. Approximately 20 GPs record each week with around 1,000 GPs participating annually.

While BEACH data fields do not change over time, there is also a small amount of space on each form in which supplementary data can be collected. This is referred to as Supplementary Analysis of Nominated Data (SAND). These may be sponsored by different interest groups for short blocks of data collection and may include information about the patient's health, management and risk factors such as smoking and alcohol consumption (Britt et al. 2001). Ten to twenty SAND topics are covered annually. Reports are made available to the organisations sponsoring the SAND. Several SAND modules have been collected on asthma.

The response rate for BEACH is fairly low. In the 2005–06 BEACH year (April 2005 to March 2006), only 31.1% of GPs who were contacted and available agreed to participate. A comparison of participating GPs and all recognised GPs in the sample frame revealed that GPs aged 55 years and older were over-represented and those younger than 35 years of age were under-represented in the BEACH participants. There is no information available on patients of GPs who refused to participate. As younger GPs see younger patients and possibly those with more acute conditions, then younger patients and certain conditions may be underestimated in the sample. Post-stratification weighting of the sample is used to adjust for these age differences and activity level of the GPs and ensure that the sample is representative.

Although there are limitations in the validity and reliability of BEACH, particularly at the detailed level, it is currently the only reasonably reliable and valid method of collecting information about morbidity and its management in general practice across Australia.

8.2 Access to BEACH data

ACAM currently has approval from the AGPSCC to access BEACH data collected between 1998 and 2005 for the purpose of reporting on general practice care of patients with asthma.

8.3 Data items

8.3.1 BEACH data

The BEACH collection includes information about the following:

The encounter

- date and type of consultation
- up to four diagnoses or problems managed
- Medicare/Veterans' Affairs item number.

The patient

- age and sex
- postcode of residence
- health care card status; Veterans' Affairs card status
- non-English-speaking background status
- whether the patient identifies as Aboriginal and/or Torres Strait Islander
- up to three reasons for the encounter.

The management of each problem

- medications prescribed, supplied or advised including brand, form, strength, and dosage
- non-pharmacological management including counselling, referrals, procedures, pathology and imaging ordered.

The GP characteristics

- age and sex
- years working in general practice
- number of sessions worked per week
- postcode of main practice, etc. (Britt et al. 2001).

Disaggregation of BEACH data is possible by age, sex, Aboriginal and Torres Strait Islander status, ethnicity (non-English-speaking background or not), socioeconomic status (based on postcode or healthcare card status), and rural and remote status (RRMA). However, the rate of GP consultations for Aboriginal and Torres Strait Islander peoples is likely to be an underestimate of the true consultation rate (AIHW & ABS 2006; AIHW GPSCU 2003). Due to the small sample size of the SAND, disaggregation may be limited to age, sex, ethnicity (non-English-speaking background or not) and socioeconomic status (based on postcode or healthcare card status).

8.3.2 SAND

The SAND are collected as a supplementary data set of the BEACH program (Britt et al. 2001). Organisations sponsoring blocks of SAND collection ask questions on topics of their choice and have access to the detailed reports. GPs participating in SAND ask and record responses to specific questions in targeted patient groups. SAND modules relevant to asthma have been conducted in 1999, 2000–01, 2002, 2004 and 2006 (Britt & Miller 2007).

8.3.3 International Classification of Primary Care

Information on diagnosis and problem managed during GP encounters, obtained from the BEACH data set, are classified according to the International Classification of Primary Care—2nd edition (ICPC-2) (Britt et al. 1998; Britt et al. 2001; WICC (Classification Committee of the World Organization of Family Doctors) 1998).

8.4 Using BEACH data for asthma monitoring

ACAM has used BEACH data to describe and monitor trends in the rate of general practice encounters for asthma in Australia, as well as patterns by age, sex, state and territory, remoteness and socioeconomic status.

For example, the estimated number of general practice encounters for asthma per 100 population is estimated as follows:

1. The number of asthma-related general practice encounters (ARGPEs) per 100 general practice encounters is estimated from the BEACH data set using a method which adjusts for the cluster (practice-based) sampling used in BEACH and also incorporates post-stratification weights to account for differences in age between the GP sample and the GP population (see also Section 2.4).
2. The estimated total number of all general practice visits is calculated from Medicare on-line data for Medical Benefits Schedule (MBS) Category 1 Service Items (see Section 10.1.1). This category includes all unreferred (that is primary care) attendances (HIC 2002).
3. The estimated number of ARGPEs per 100 population is then estimated as:

$$\frac{\text{ARGPEs per 100 general practice encounters} \times \text{estimated total number of all general practice visits}}{\text{population}}$$

where population = the mid-year population for the relevant year(s).

The 95% confidence interval for the number of general practice encounters per 100 population is estimated using the 95% confidence interval for the ARGPEs per 100 GP encounters in the above equation.

9 Pharmaceutical Benefits Scheme data

9.1 Background

The Pharmaceutical Benefits Scheme (PBS) is an Australian Government initiative that provides subsidy for prescribed drugs dispensed to all Australians and to eligible overseas visitors. This scheme has been in existence for 60 years. Recent changes in the data recorded within the PBS data set have enabled the anonymous identification of individuals within the data set linked to basic demographic details with the inclusion of the patient Medicare number. These data may be used to describe patterns of medication use in the population.

Data from the PBS are provided to ACAM by the Pricing and Estimates Section of the Department of Health and Ageing. A letter of request should be addressed to them, with a list of the variables and the item numbers of the medications needed for analysis. The data request should specify the time period for which data are required. In practice, the earliest date for which data linked to Medicare numbers are available is May 2002. The request should list the PBS item codes of the drugs of interest and the demographic data required. The demographic data will not be provided in a manner that is potentially identifying. Hence, age group will be provided but not actual age or date of birth.

9.2 Data coverage

The PBS data set contains records for all prescriptions for which a subsidy was paid by the PBS. There are two large groups of medications dispensed that are excluded from this data set:

- Medications that are purchased without a prescription (over-the-counter), for example, short-acting bronchodilators, simple analgesics etc.
- Medications that cost less than the co-payment threshold. The co-payment threshold varies both between and within individuals. In general, individuals who hold a government concession card or who have reached the 'safety net' threshold within a calendar year are required to pay a lower, concessional co-payment. Individuals who do not meet these criteria pay the full co-payment. The amounts vary from year to year but are currently around $5 and $30, respectively.

Many cheaper prescription medications, such as oral corticosteroids, fall below the general co-payment threshold but above the concessional co-payment threshold. Data on medications that fall within this range are available for concessional patients but not for general patients.

Some over-the-counter medications, such as short-acting bronchodilators, are also available on prescription and are cheaper for concessional patients to purchase on prescription. Therefore, we can assume that most concessional patients will have purchased short-acting bronchodilators with a subsidised (and, hence, recorded) prescription but most general patients will have purchased the drug over-the-counter and hence not be recorded.

For these reasons, analyses of data for some classes of medications is limited to data from concessional patients.

There is lag in the processing of prescription data so that the date of processing data may be up to 3 months after the actual date of supply. For example, if the date of processing is up to the end of December 2006 records with date of supply up to September 2006 will be about 99.6% complete. Table 9.1 shows the completeness of PBS data, in percent, based on date of supply when date of processing is December 2006.

Table 9.1: Completeness of PBS data based on date of supply

Date of supply	Percent complete
September 2006	99.6
October 2006	99.1
November 2006	95.7
December 2006	24.7

Note: Date of processing December 2006.

9.3 Data items

Each record of the PBS data refers to a single episode of supply of a single item, identified by its PBS item number and the date of supply. However, occasionally more than one unit of a single item may be supplied at the same time. This is recorded in data field ('Scripts'). Individual recipients are anonymously identified within the data set by a scrambled Personal Identification Number ('PIN'). The record also contains information on the patient category (concessional or general patient), sex, age group, and residential postcode of the recipient.

For the analysis of PBS data undertaken by ACAM in 2005–06 for the purposes of the *Patterns of asthma medication use in Australia* report (ACAM 2007a), the original data set included the variables shown in Table 9.2.

The data set contained a total of 19,181,628 records. To cope with this large number of records, computing power was increased by upgrading the analysis computer and enhancing memory capacity (upgraded to Pentium 4 Computer, 3.20GHz with 3.0GB RAM). To lessen the traffic on the network, the data was temporarily saved onto the local hard drive.

Table 9.2: Pharmaceutical Benefits Scheme variables obtained from the Pricing and Estimates Section of the Department of Health and Ageing

Variable name	Description	Data categories	Type and length of variable
PIN	Encrypted personal identification number that uniquely identified individuals in the data set	Unique to each individual	Numeric, 9 spaces
Dtsupply	Date of supply	N/A	Character, 10 spaces
Sex	Gender of patient	Male, female	Character, 1 space
Pcode	Residential postcode of patient at date of supply	N/A	Numeric, 4 spaces
Pcatgry	Patient concession category at date of supply	General safety net, general non-safety net, concessional safety net, concessional non-safety net, repatriation PBS concession card holder safety and non-safety net	Character, 2 spaces
Agegr	Five year age group of patient at date of supply	Age 0 to 4 years through to 85 years and over	Character, 5 spaces
Item	Item number of drug dispensed	Items for all drugs classified as: inhaled corticosteroids, long-acting beta agonists, short-acting beta agonists and oral corticosteroids	Character, 5 spaces
Scripts	Number of prescriptions dispensed for each item for each date of supply per patient	N/A	Numeric, 12 spaces

9.4 Data processing

Data checking, cleaning and analyses can be carried out using any statistical software that can handle large data sets. Before any analyses are undertaken, it is important to take the time to get to know the data. The following steps are recommended in processing the data:

1. Import the raw PBS data file into a data file format that is readable by the statistical software of choice.
2. Check the frequency distribution of each variable to identify outliers and invalid records (for example medication item numbers that are not a drug of interest). Having examined the distributions, it is possible to make decisions on which outliers are most likely to be errors.
3. Allocate PINs to records that cannot be matched due to incorrect personal details or incorrect Medicare number. These 'dummy PINs' can be identified in a separate data set provided by the Pricing and Estimates Section of the Department of Health and Ageing. This file is merged with the original data set by PIN to identify dummy PIN records. All records for dummy PINS are excluded, leaving a new data set in which all the PINs are 'real'.
4. Limit the data set to the pre-defined study period by excluding any records with dates of supply outside that period.

5. Identify individuals within the data set.
    a. Create separate data sets (4) for all records in each drug class (inhaled corticosteroids, oral corticosteroids, long-acting beta agonists, short-acting beta agonists).
    b. Summarise each drug class data set to determine the total number of prescriptions per person (identified using the PIN)
    c. Merge the summarised data sets back together by PIN. The output will be a table in which there is one record (row) per PIN with (variables) columns containing the number of prescriptions dispensed for each drug class.
6. Merge this file with the original file containing postcodes and other demographic variables, by PIN to create the analysis data set.
7. Merge the above file with a file assigning SEIFA and ASGC categories to the residential postcodes*. Where there is more than one postcode associated with a PIN, the postcode from the earliest record should be used.
8. Delete records not related to asthma medications. ACAM wished to limit the analysis to medications that had been supplied for the management of asthma (or at least obstructive lung disease). There are no diagnosis details available within the PBS data. However, we assumed that inhaled medications (inhaled long and short-acting beta agonists and inhaled corticosteroids) were supplied for the management of obstructive lung disease. However, oral corticosteroids are used to treat many non-respiratory conditions as well as obstructive lung disease. Therefore, it is important to exclude individuals who were only dispensed oral corticosteroids and no other asthma medications to increase the likelihood that the oral corticosteroid prescriptions in the data set were used for obstructive airways disease (asthma and COPD). This can be achieved by deleting records from the above file that contain one or more prescriptions for oral corticosteroids and none for the other asthma drug classes. The resulting file is labelled the 'asthmapin' file.

* In the case of ASGC, postcodes often match to more than one ASGC category because a postcode can be shared by two or three different categories where proportions of the postcode cover different ASCG categories. For example, postcode 0822 falls into three ASGC categories, *Outer Regional, Remote and Very Remote* Australia, with 22.82%, 6.64% and 70.54%, respectively. In these cases when using SAS the RANTBL function can be used to generate a random number from the probability distribution of the proportion of the postcode that falls into each ASGC category. The random number generated indicates which of the ASGC categories to assign the postcode to. Other statistical packages will have a corresponding facility that generates random numbers.

9.5 Data analyses

Many drugs within the same class have different potencies. To be able to combine information from several drugs within a class, it is necessary to convert them to the same unit of measurement. Furthermore, when comparing drug use between classes of drugs, it is helpful to have a common unit of measurement which reflects the standard or common usage of each class. The World Health Organization (WHO) has published a common

currency or unit of drug consumption known as defined daily doses (DDDs) (WHO (World Health Organization) Collaborating Centre for Drug Statistics Methodology 2006).

The DDD is a standard unit of measurement that enables the comparison of the rate of use of various drugs in individuals and in populations. It is defined as the assumed average maintenance dose per day for a drug used for its main indication in adults.

The rate of use of drugs in individuals is reported as the DDD per person per day:

1) $DDD\ per\ person\ per\ day = \frac{N * M * Q}{DDD * Y}$

The rate of use of drugs in populations is reported as the DDD per 1,000 population per day:

2) $DDD\ per\ 1{,}000\ population\ \text{per}\ day = \frac{N * M * Q * 1{,}000}{DDD * P * Y}$

where

N is the number of prescriptions dispensed per record

M is the strength of each dose (milligrams or micrograms)

Q is the average quantity of doses per prescription

DDD unit is one defined daily dose for the particular item classified by the World Health Organization Collaborating Centre for Drug Statistics Methodology

Y is the number of days within the study period

P is the population of interest.

For example, a person was given 24 prescriptions of salbutamol nebuliser solution for 1 year. Salbutamol has a DDD unit of 10mg. The dispensed formulation of salbutamol had 2.5mg of the active ingredient per dose and one prescription included 60 doses. Using formula number 1, the person's average maintenance dose of salbutamol per day (in DDD per person per day) is:

$$DDD\ per\ person\ per\ day = \frac{24\ prescriptions * 2.5\ mg\ strength\ per\ dose * 60\ doses\ per\ prescription}{10\ mg\ DDD\ unit * 365\ days}$$

= 0.98 DDD of salbutamol per person per day.

In a certain area, persons were dispensed a total of 2,400 prescriptions of salbutamol nebuliser solution (2.5mg per dose, 60 doses per prescription) in a year. If the population in that area was 10,000, we use formula number 2, and we can do it in two steps:

1) $DDD\ per\ item = \frac{2{,}400\ prescriptions * 2.5\ mg\ strength\ per\ dose * 60\ doses\ per\ prescription}{10\ mg\ DDD\ unit}$

= 36,000 DDD per item

2) $DDD\ per\ 1{,}000\ persons\ per\ day = \frac{36{,}000\ DDD\ per\ item}{10{,}000\ persons * 365\ days} * 1{,}000$

= 9.86 DDD of salbutamol per 1,000 persons per day.

In fact, there are other drug items for salbutamol (for example, a 5mg nebuliser solution as well as metered dose inhaler formulations). Therefore, it is necessary to calculate the DDD for each salbutamol item, and then add all DDDs per item to get the total for salbutamol before dividing it by the product of the population and the number of days. Multiply the result by 1,000 to get the rate.

The following is a step-by-step procedure to calculate the rate of use of medication using PBS data.

9.5.1 Calculation of the DDD per person per day for each drug class

1. Using the analysis data set, make separate datasets for each drug class (four data sets). In our study, the data sets for short-acting beta agonists and oral corticosteroids were restricted to patients aged 15 years and over that were concession card holders (see page 50).
2. Calculate the total number of prescriptions per drug item number per PIN.
3. Create another four datasets for each drug class, containing one record for each drug item within that class. Each record contains fields for the strength per dose (in milligrams or micrograms), the dispensed quantity of doses per prescription, and the DDD unit of each drug class. This can be called the 'item dose data set' for a particular drug class.
4. Merge the 'Item dose data set' with the analysis dataset by item number.
5. A person can have different dates of supply, depending on how many times their prescriptions were registered with the PBS. Extract only the first date of supply of any medication in that class. Calculate the number of days from the first date of supply to the end of the study period (in our example, 1st of July 2004). Those with less than 7 days since first supply were excluded to reduce inflation of the DDD per person per day.
6. The resulting data sets will now have the PIN, the drug item number, the total number of prescriptions (N), the corresponding strength (M), the average dose (Q), and the DDD unit (DDD) per drug item, plus the number of days within the study period (Y).
7. Calculate the DDD units for each item for each person as:
    a. *DDD per item per person = N*M*Q/DDD*
    b. In this step, records with a total number of prescriptions (N) for a single item that are more than 96 were excluded, as these outliers are likely to be errors and could be influential on the final estimates.
8. After calculating the DDD per item per person, sum all the DDDs for items within an asthma drug class to get the total DDD per drug class for each PIN.
9. Obtain the average rate of use of a particular asthma medication class per person by dividing the total DDD in each PIN by the total number of days between the person's first date of supply of an item in the class to the end of the period of observation (Y):

    DDD per person per day = Total DDD per person/total number of days.

10. Chart the DDD per person per day as a relative frequency histogram and a cumulative relative frequency line plot:
    a. As the data are highly skewed to the right, the graph is more interpretable if the values of DDD per person per day are log10-transformed.
    b. If using Excel, choose the 'Custom types – Line column on 2 axes', with the 'percent' as the first Y axis and the 'cumulative percent' as the second Y axis. The labels of the X axis should be the original values in DDD per person per day, not the log transformed variable.
    c. It is useful to mark the median and interquartile range of the distribution. This can be done using the cumulative relative frequency plot.

9.5.2 Calculation of the DDD per 1,000 population per day

This is a measure of population level usage of medications. Formula number 2 will be used to calculate this rate. The only difference with formula number 1 is the inclusion of the population in the denominator and multiplying by 1,000, to get the rate per 1,000 population. The number of days (Y) is the duration of the study period (that is 731 days for the period 1 July 2002 to 30 June 2004).

The above calculations can be done in two steps:

- Calculate the total number of DDDs within the relevant drug class dispensed to the population of interest during the study period as described above.
- Divide the total DDDs in a drug class by the total number of persons in the population of interest, and then multiply by 1,000.

Denominator populations (P)

For inhaled corticosteroids and long-acting beta agonists, in which the full data set could be included, the denominator population is the Australian estimated resident population in the relevant year. For short-acting beta agonists and oral corticosteroids, where analyses are limited to government health card holders, the denominator population is the estimated total number of government health concession card holders in Australia. This estimate is obtained from the National Health Survey data (using the NHS CURF).

When doing analyses in subgroups of the population (for example specific age-groups) use the population of the relevant subgroup as the denominator.

10 Medicare Australia Online data

10.1 Medical Benefits Scheme data

Medicare Australia provides statistics on the claims made to the MBS. These include items claimed for services performed by general practitioners, doctors and specialists. It can be used to obtain data on GP consultations, claims made under the Asthma Cycle of Care Practice Incentive Program (PIP) and for procedures, such as spirometry. Online interactive data reports for MBS items can be created from January 1993 to within 3 months of the current point in time and are available at:

<www.medicareaustralia.gov.au/statistics/dyn_mbs/forms/mbs_tab4.shtml>.

Reports on individual benefit item numbers can be collated by time period, state, age and sex. More complex reports are available directly from Medicare Australia on a fee for service basis.

The MBS data only include information on items claimed by general practitioners, doctors and specialists in the community. The principal items that may be of use for monitoring asthma indicators are spirometry and the Asthma Cycle of Care Service Incentive Payment (SIP) items. There is no diagnosis information attached to the MBS items, therefore spirometry items may relate to conditions other than asthma, notably COPD or screening in normal individuals. Information about spirometry or asthma management in public hospital or outpatient settings is not available in the MBS data.

There is no published information on the quality of MBS statistics. However, as the data are based on Medicare details and claims made by medical practitioners or patients for reimbursement they are likely to provide quite accurate information.

The utility of the MBS data for disease monitoring would be greatly enhanced if it was possible to link these data with other data such as PBS, hospitalisation and/or mortality data sets.

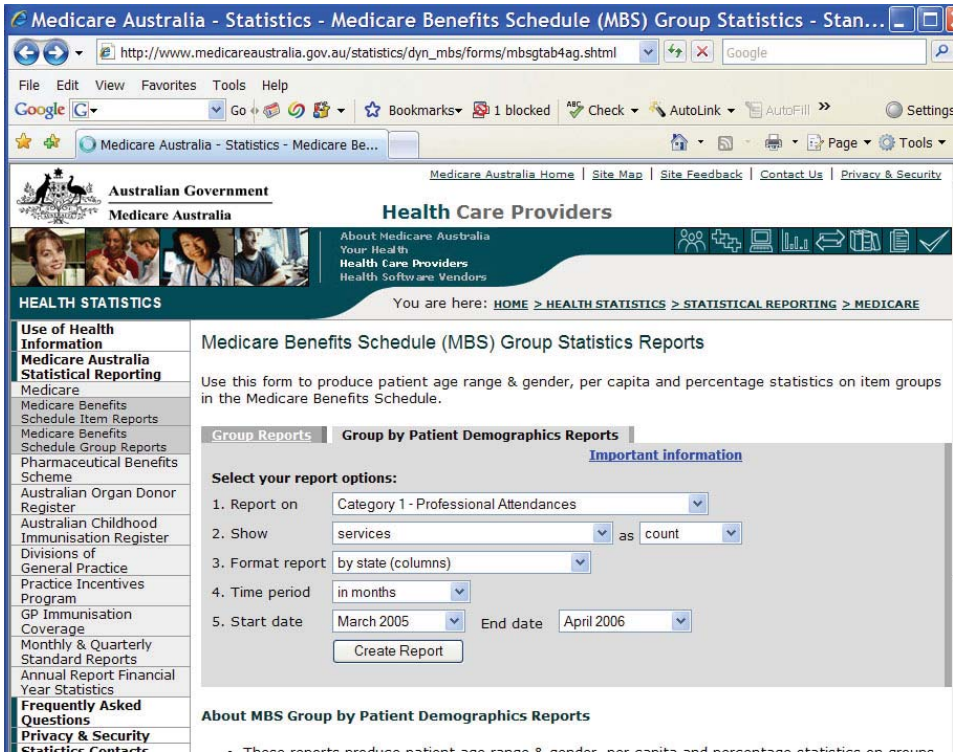
10.1.1 Data on GP attendances

For the purposes of analysing BEACH data, it is necessary to obtain the total number of GP attendances in a given time period as the denominator population.

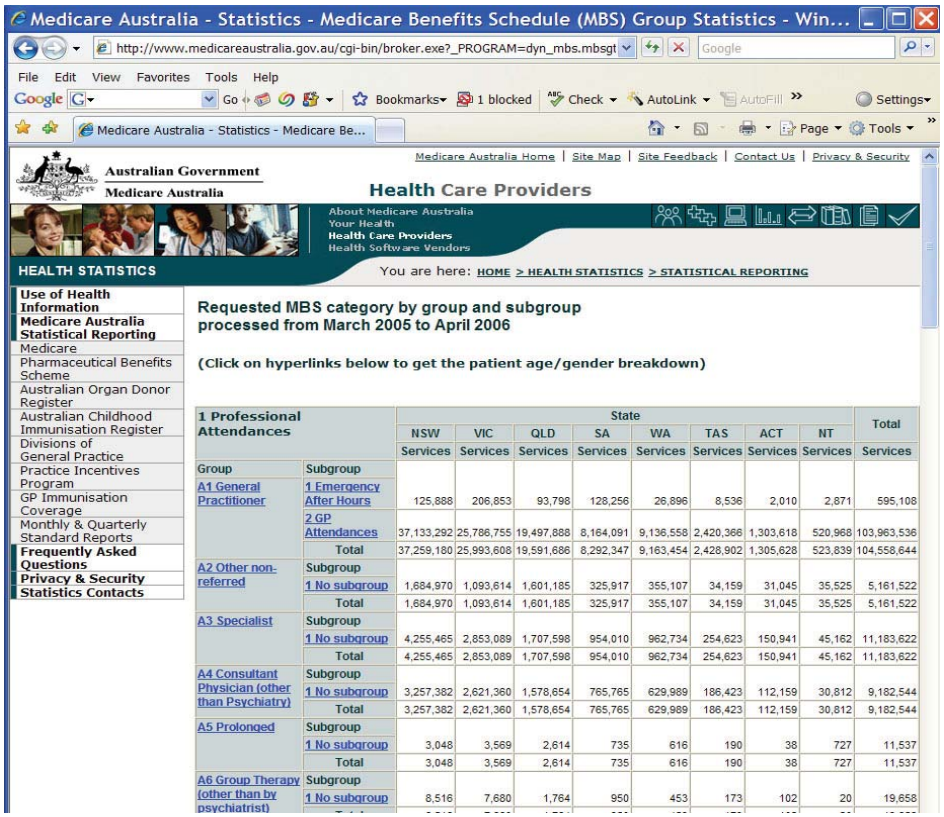
The number of GP attendances in a given time period, and by state/territory, and age and sex, can be accessed at:

<www.medicareaustralia.gov.au/statistics/dyn_mbs/forms/mbsgtab4ag.shtml>.

and by selecting the following report options:



The following report will be produced:



1 Professional Attendances		State								Total
		NSW	VIC	QLD	SA	WA	TAS	ACT	NT	
		Services	Services	Services	Services	Services	Services	Services	Services	Services
Group	Subgroup									
A1 General Practitioner	1 Emergency After Hours	125,888	206,853	93,798	128,256	26,896	8,536	2,010	2,871	595,108
	2 GP Attendances	37,133,292	25,786,755	19,497,888	8,164,091	9,136,558	2,420,366	1,303,618	520,968	103,963,536
	Total	37,259,180	25,993,608	19,591,686	8,292,347	9,163,454	2,428,902	1,305,628	523,839	104,558,644
A2 Other non-referred	Subgroup									
	1 No subgroup	1,684,970	1,093,614	1,601,185	325,917	355,107	34,159	31,045	35,525	5,161,522
	Total	1,684,970	1,093,614	1,601,185	325,917	355,107	34,159	31,045	35,525	5,161,522
A3 Specialist	Subgroup									
	1 No subgroup	4,255,465	2,853,089	1,707,598	954,010	962,734	254,623	150,941	45,162	11,183,622
	Total	4,255,465	2,853,089	1,707,598	954,010	962,734	254,623	150,941	45,162	11,183,622
A4 Consultant Physician (other than Psychiatry)	Subgroup									
	1 No subgroup	3,257,382	2,621,360	1,578,654	765,765	629,989	186,423	112,159	30,812	9,182,544
	Total	3,257,382	2,621,360	1,578,654	765,765	629,989	186,423	112,159	30,812	9,182,544
A5 Prolonged	Subgroup									
	1 No subgroup	3,048	3,569	2,614	735	616	190	38	727	11,537
	Total	3,048	3,569	2,614	735	616	190	38	727	11,537
A6 Group Therapy (other than by psychiatrist)	Subgroup									
	1 No subgroup	8,516	7,680	1,764	950	453	173	102	20	19,658

When calculating the total number of GP attendances in a given time period, not all professional attendances are included. We only include those that fall under the categories of:

- A1—general practitioner (including both subgroups of 1—Emergency after hours and 2—GP attendances)
- A2—other non-referred.

It is best to only download six months data at a time as downloading data for a longer period of time is very slow and can sometimes create errors.

10.2 Pharmaceutical Benefits Scheme online

In addition to our own analyses of PBS record level data (described in Section 9, above), ACAM has utilised online interactive reports from both the PBS and the Repatriation Pharmaceutical Benefits Scheme (RPBS), which are available from Medicare Australia. Online interactive data reports for PBS statistics can be created from January 1992 at:

<www.medicareaustralia.gov.au/statistics/dyn_pbs/forms/pbsgtab1.shtml>.

Reports can be collated for selected pharmaceutical benefits item numbers or by therapeutic grouping. They can be categorised by state, patient category (General, Concessional, Repatriation PBS), Anatomical Therapeutic Chemical (ATC) Classification and time period.

11 Other data sources

11.1 Pharmacy Guild Survey

The Pharmacy Guild Survey is an ongoing survey established in 1990, which collects total dispensing information for each year from a stratified random sample of 150–250 pharmacies. The sample was selected so that it is representative of all operational pharmacies based on PBS dispensing volume and geographical location. From the survey data provides estimates of the prescription volumes for medicines that are not subsidised by the PBS. The volumes of each non-subsidised drug are estimated by multiplying the total survey estimates by a weighting factor.

Data from the Pharmacy Guild Survey are available in volumes of *Australian statistics on medicines*, published by the Department of Health and Ageing. The Department of Health and Ageing combines the actual counts of those prescription categories available through the PBS data with the prescription estimates for the non-subsidised community prescriptions from the Pharmacy Guild Survey. The Pharmacy Guild Survey does not collect any information on patient demographics or the condition for which the medication is being prescribed, nor does it collect enough information to allow disaggregation at any level.

11.2 IMS Health data

IMS Health is a private company that collects data on the sales of pharmaceutical products on behalf of major manufacturers and wholesalers operating in Australia and worldwide.

The value of these data is that they reflect supply (and, hence, purchases) of specific medications. As some medications are sold without prescription or are below the subsidy threshold, equivalent data are not available through the PBS.

ACAM has calculated the annual aggregate number of packs distributed each year for each product relevant to the treatment of asthma for the period January 1996 to December 2004. Parenteral forms are excluded. Usage is measured in DDDs/1,000 persons/day as described in Chapter 9 (PBS data).

The main limitation of the data is that there is no information on the characteristics of the purchasers or consumers. As these drugs are commonly used for people with COPD, it is not possible to directly ascribe the trends and differentials observed in these data to the population with asthma. Furthermore, socioeconomic and geographic trends and differentials in the use of drugs cannot be assessed using these data.

Appendix 1 National asthma indicators

Table A.1: Asthma indicators and their data sources

Indicator	Description	Data source(s)	Data custodian(s)
Prevalence			
Prevalence of ever having doctor-diagnosed asthma	The prevalence rate of ever having doctor-diagnosed asthma per 100,000 resident population	National Health Survey State/territory CATI surveys	ABS Relevant state/territory health departments
Prevalence of current asthma	Prevalence rate of people ever diagnosed with asthma who have experienced symptoms of asthma (wheeze, shortness of breath or chest tightness) or taken treatment for asthma in the last 12 months per 100,000 resident population	National Health Survey * State/territory CATI surveys	ABS Relevant state/territory health departments
Prevalence of wheeze in the preceding 12 months	Prevalence rate of wheeze or whistling in the chest in the previous 12 months per 100,000 resident population	No national data available at this time	
Mortality			
Death rate for asthma, ages 5 to 34 years	Death rate in people aged 5 to 34 years where the underlying cause of death is asthma (ICD-10-AM codes: J45, J46) per 100,000 resident population	National Mortality Data Collection	AIHW
Death rate for asthma, all ages	Death rate in people all ages where the underlying cause of death is asthma (ICD-10-AM codes: J45, J46) per 100,000 resident population	National Mortality Data Collection	AIHW
Health care utilisation			
Hospital separations for asthma	Hospital separation rate for asthma (ICD-10-AM codes: J45, J46) per 100,000 resident population	National Hospital Morbidity Database	AIHW
Hospital patient days for asthma	Hospital patient days attributable to asthma (ICD-10-AM codes: J45, J46) per 100,000 resident population per year	National Hospital Morbidity Database	AIHW
Individuals with one or more hospital separations for asthma	Rate of having one or more hospital separations for asthma in a year per 1,000 resident population	Data linkage using National Hospital Morbidity Database	AIHW
Hospital re-admissions for asthma	Hospital re-admission rate for asthma per 1,000 resident population	Probabilistic linkage using National Hospital Morbidity Database	AIHW
Hospital re-attendance for asthma within 28 days	Number of people discharged with a principal diagnosis of asthma (ICD-10-AM J45 or J46) who re-attend at an emergency department (ED) and/or are re-admitted to hospital within 28 days with a diagnosis of asthma	Probabilistic linkage using National Hospital Morbidity Database	AIHW
Asthma-related ED attendance	Rate of hospital ED attendances for asthma per 100,000 resident population	No national data source available at this time, however data from NSW and Victoria have been used	NSW Department of Health, Victorian Department of Human Services
Rate of asthma-related general practice encounters	Rate of asthma-related general practice encounters per 100 resident population	BEACH data and Medicare Australia statistics	Australian General Practice Statistics and Classification Centre (AGPSCC) and DoHA

(continued)

Table A.1 (continued): Asthma indicators and their data sources

Indicator	Description	Data source(s)	Data custodian(s)
Management			
Proportion of people with asthma who have an asthma action plan	Proportion of people with current asthma who have an individualised, written asthma action plan with the four essential components, developed in consultation with a health professional, per 100,000 resident population	National Health Survey * State/territory CATI surveys	ABS Relevant state/territory health departments
Proportion of people with asthma who use preventers regularly	Proportion of people with asthma who use a preventer medication regularly	National Health Survey * State/territory CATI surveys	ABS Relevant state/territory health departments
Quality of life			
Impact of asthma on quality of life	Proportion of people with asthma who report poor health-related quality of life	National Health Survey * State/territory CATI surveys	ABS Relevant state/territory health departments
Smoking and exposure to passive smoke among people with asthma			
Rate of smoking in people with asthma	The proportion of people aged 18 years or over who have current asthma and who smoke any tobacco product weekly or more frequently per 100,000 resident population	National Health Survey * State/territory CATI surveys	ABS Relevant state/territory health departments
Prevalence of smoking in households where children with asthma reside	The proportion of children aged under 15 years who live in a household where people smoke and who (a) have current asthma or (b) have had wheeze in the previous 12 months	National Health Survey * State/territory CATI surveys	ABS Relevant state/territory health departments
Other asthma indicators			
Prevalence of airway hyperresponsiveness		No national data available at this time	
Proportion of people with asthma who have had recent spirometry		Medicare Australia	DoHA
Index of asthma control	Composite indicator	No national data available at this time. Some relevant elements are available through state/ territory CATI surveys	
Incidence rate of asthma initiated (caused) by occupational exposure		No national data available at this time	
Rate of Asthma Cycle of Care (formerly Asthma 3+ Visit Plan) Practitioner Incentive Program payments	Rate of payments for completed Asthma Cycle of Care Practitioner Incentive Program payments per 100,000 resident population	Medicare Australia statistics	DoHA
Proportion of schools using the Asthma Friendly Schools Program	Proportion of schools recognised as using the Asthma Friendly Schools Program		State and territory Asthma Foundations
Costs of asthma to individuals	Composite indicator	No national data available at this time	

* data source provides some information on the relevant asthma indicator, but does not perfectly align with the recommended ACAM definition of that indicator

Glossary

Aboriginal	A person of Aboriginal descent who identifies as an Aboriginal and is accepted as such by the community in which he or she lives.
Admission	Admission to hospital. In this report, the number of separations has been taken as the number of admissions. Hence, admission rate is the same as separation rate.
Age-specific rate	A rate for a specified age group. The numerator and denominator relate to the same age group. See Section 2 for full description.
Age-standardisation	A method of removing the influence of age when comparing populations with different age structures. This is usually necessary because the rates of many diseases vary considerably with age. The age structures of the different populations are converted to the same 'standard' structure, then the disease rates that would have occurred with that structure are calculated and compared. See Section 2 for full description.
ARIA/ASGC classification	A classification of the level of accessibility to goods and services (such as general practitioners, hospitals and specialist care) based on the proximity to these services (measured by road distance).
Associated cause of death	Diseases, conditions or injuries that contributed to the death directly or indirectly. Compare with *Underlying cause of death*.
Asthma	A chronic inflammatory disorder of the airways in which many cells and cellular elements play a role, in particular mast cells, eosinophils, T lymphocytes, macrophages, neutrophils and epithelial cells. In susceptible individuals this inflammation causes recurrent episodes of wheezing, breathlessness, chest tightness and coughing, particularly at night or in the early morning. These episodes are usually associated with widespread but variable airflow obstruction that is often reversible either spontaneously or with treatment. The inflammation also causes increases in existing bronchial hyperresponsiveness to a variety of stimuli. (NAEPP (National Asthma Education and Prevention Program) 1997)
Asthma action plan	A plan that provides instructions on how to recognise and respond to worsening asthma. It is recommended that these instructions be given in writing ('written asthma action plan'). The action plan is based on symptoms and/or peak expiratory flow measurements and is individualised according to the pattern of the person's asthma. These plans have sometimes been referred to as 'asthma management plans', 'asthma plans', 'self-management plans', 'asthma care plans' or 'personal asthma plans'.

Asthma management plan	An individualised plan of management for patients with asthma formulated in accordance with the Six Step Asthma Management Plan. (The asthma action plan forms one part of this.)
Asthma 3+ Visit Plan/Asthma Cycle of Care	An incentive scheme funded by the Australian Government aimed at people with moderate to severe asthma. The plan entails three visits to the GP at which asthma is assessed, an individualised asthma management plan is developed and reviewed, and the patient receives appropriate education about asthma.
Average length of stay (ALOS)	The average of the length of stay for admitted patient episodes. Calculated by dividing total patient days in a given period by the total number of hospital separations in that period. See *Patient days, Hospital separation* and *Length of stay.*
BEACH survey	A continuous cross-sectional paper-based data collection, which collects information about the reasons for seeking medical care, the type of patients seen, the types of problems managed and treatment provided in general practice across Australia.
Cause of death	The disease or factor contributing to the death. When used technically this term is usually applied to the 'underlying cause' listed on the medical certificate issued at death. From information reported on the medical certificate of cause of death, each death is classified by the underlying cause of death according to rules and conventions of the International Classification of Diseases of the day (currently ICD version 10). See *Underlying cause of death* and *Associated cause of death.*
Confidence interval	A statistical term describing a range (interval) of values within which we can be 'confident' that the true value lies. For example, a 95% confidence interval implies that there is 95% confidence that the true value will be included in this interval.
Crude death rate	The number of deaths in a given period divided by the size of the corresponding population indexed to 100,000. See Section 2 for full description.
Defined daily dose	The assumed average maintenance dose per day for a drug used for its main indication in adults.
Disability-adjusted life year (DALY)	Years of healthy life lost through premature death or living with disability due to illness or injury.
Estimated resident population	An estimate of the resident population derived from the 5-yearly Census counts. It is based on the usual residence of the person.
Health-related quality of life	A term used to describe the impact that a disease has on an individual's health status and everyday functioning. It is most often used when referring to chronic diseases.

Health risk factor	Any factor that represents a greater risk of a health disorder or other unwanted condition. Risk factors may be causes of disease or contributors to disease.
Health service use	Use of the available health care services within the population, including hospitals, emergency departments and general practitioners.
Health survey	A research method in which health information is collected from participants at a point in time. In population health monitoring, this typically involves selecting a representative sample of the population and administering questionnaires to the participants. This can be done in person, over the phone or by post. Some surveys have additionally included physiological measurements.
Hospital separation	The formal process by which a hospital records the completion of treatment and/or care for an admitted patient. The episode of care may be completed by an admitted patient's discharge, death, transfer to another hospital or change in the type of care.
Incidence	The number of new cases (of a disease, condition or event) occurring during a given period. Compare with *Prevalence.*
Indicator	A key statistic chosen to describe (indicate) a situation concisely, help assess progress and performance, and act as a guide to decision making. It may have an indirect meaning as well as a direct one; for example, Australia's overall death rate is a direct measure of mortality but is often used as a major indicator of population health.
Indigenous Australians	Refers to people who identify themselves as being of Aboriginal or Torres Strait Islander origin.
International Classification of Diseases (ICD)	The World Health Organization's internationally accepted statistical classification of disease and injury. The 10th revision is currently in use.
Length of stay	Duration of hospital stay, calculated by subtracting the date the patient is admitted from the day of separation. All leave days, including the day the patient went on leave, are excluded. See also *Average length of stay.*
Medicare	A national, government-funded scheme that subsidises the cost of personal medical services for all Australians and aims to help them afford medical care.
Morbidity	Refers to ill-health in an individual and to levels of ill-health in a population or group.
Mortality	Death.

Outcome (health outcome)	A health-related change due to a preventive or clinical intervention or service. (The intervention may be single or multiple and the outcome may relate to a person, group or population or be partly or wholly due to the intervention.)
p-value	The probability that the observed difference or association could have occurred by chance. If that probability is less than 5% (0.05), it is conventionally held that it did not occur by chance and is a true difference or association.
Patient days	The total number of days for patients who were admitted to hospital for an episode of care and who separated during a specified reference period. A patient who is admitted and separated on the same day is allocated one patient day. Compare with *Length of stay* and *Average length of stay*.
Pharmaceutical Benefits Scheme (PBS)	A national, government-funded scheme that subsidises the cost of a wide range of pharmaceutical drugs, and that covers all Australians to help them afford standard medications.
Prescription drugs	Pharmaceutical drugs available only on the prescription of a registered medical practitioner and only from pharmacies.
Prevalence	The number or proportion of people with certain conditions in a population at a given time. Compare with *Incidence.*
Principal diagnosis	The diagnosis established to be chiefly responsible for occasioning the episode of care or attendance at a health care facility.
Quintile	A group derived by ranking the population according to specified criteria and dividing it into five equal parts.
Re-admission	An admission to the same or different hospital within a defined period following discharge from a hospital. Transfers to other hospitals and changes in care type are not included as re-admissions.
Re-attendance	An admission to a hospital or visit to the emergency department within a defined period following discharge from hospital or the emergency department.
Risk factor	See *Health risk factor*.
Rural, Remote and Metropolitan Area (RRMA)	The Rural, Remote and Metropolitan Area (RRMA) classification system divides Australia's states and territories into two metropolitan, three rural and two remote zones based on location.
Same day patient	Admitted patients who are admitted and separated on the same day.
SAND	Additional questions asked of patients in subsamples of general practice encounters, as part of the BEACH survey.

SEIFA Index of Relative Socioeconomic Disadvantage	An index of socioeconomic status which provides a summary score for a range of key socioeconomic variables that are related to health status, including household income and resources, education, occupation, fluency in English, and Indigenous status.
Separation	See *Hospital separation*.
SF-36	Short-Form 36, a widely used questionnaire containing 36 questions that measure general health and wellbeing.
Six Step Asthma Management Plan	Consensus-based guidelines for the management of asthma. The six steps are: (1) Assess asthma severity; (2) Achieve best lung function; (3) Maintain best lung function: identify and avoid trigger factors; (4) Maintain best lung function: optimise medication program; (5) Develop an action plan; and (6) Educate and review regularly.
Spirometer/spirometry	Spirometry is a measure of lung function performed by a spirometer. Spirometry is used to establish the presence of airflow obstruction and its reversibility in response to bronchodilators. This is an important feature in the diagnosis of asthma.
Statistical significance	An indication from a statistical test that an observed difference or association may be significant, or 'real', because it is unlikely to be due to chance alone. A statistical result is often said to be 'significant' if it occurs by chance only once in 20 times or less often.
Torres Strait Islander	A person of Torres Strait Islander descent who identifies as a Torres Strait Islander and is accepted as such by the community in which he or she lives.
Underlying cause of death	The condition, disease or injury initiating the sequence of events leading directly to death; that is, the primary, chief, main or principal cause. Compare with *Associated cause of death*.
Wheeze	Breathing difficulty accompanied by an audible whistling sound.

References

ABS (Australian Bureau of Statistics) 2001a. ABS views on remoteness. ABS cat. no. 1244.0. Canberra: ABS.

ABS 2001b. Demography themes: release of 1997 and 1998 ICD–10 coded mortality data. Canberra: ABS.

ABS 2002a. 2001 National health survey: summary of results. ABS cat. no. 4364.0. Canberra: ABS.

ABS 2002b. Causes of death: Australia 2001. ABS cat. no. 3303.0. Canberra: ABS.

ABS 2003a. Information paper: 2001 Census of Population and Housing–Socio-Economic Indexes for Areas. ABS cat. no. 2039.0. Canberra: ABS.

ABS 2003b. Multiple causes of death analysis. ABS cat. no. 3319.0. Canberra: ABS.

ABS 2003c. National Health Survey (Indigenous): Expanded Confidentialised Unit Record File Australia 2001 ABS cat. no. 4715.0.55.002. Canberra: ABS.

ABS 2004. Technical Paper Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA) Australia 2001. ABS cat. no. 2039.0.55.001. Canberra: ABS.

ABS 2006a. 2004–05 National Health Survey: users' guide. ABS cat. no. 4363.0.55.001. Canberra: ABS.

ABS 2006b. National Aboriginal and Torres Strait Islander Health Survey Australia 2004–05. Cat. no. 4715.0. Canberra: ABS.

ACAM (Australian Centre for Asthma Monitoring) 2003. Asthma in Australia 2003. AIHW Asthma Series 1. Cat. no. ACM 1. Canberra: AIHW.

ACAM 2005. Asthma in Australia 2005. AIHW Asthma Series 2. Cat. no. ACM 6. Available at <www.asthmamonitoring.org>. Canberra: AIHW.

ACAM 2007a. Patterns of asthma medication use in Australia. Cat. no. ACM 11. Available at <www.asthmamonitoring.org>. Canberra: AIHW.

ACAM 2007b. Survey questions for monitoring national asthma indicators. Cat. no. ACM 9. Canberra: AIHW. Viewed 9 May 2007, <www.asthmamonitoring.org>.

AIHW (Australian Institute of Health and Welfare) 2006. Australia's Health 2006. Cat. no. AUS73. Canberra: AIHW.

AIHW 2007. Australian hospital statistics 2005–06. Health services series no. 30. Cat. no. HSE 50. Canberra: AIHW.

AIHW and ABS 2006. Recent developments in the collection of Aboriginal and Torres Strait Islander health and welfare statistics 2005. AIHW cat no. IHW 15; ABS cat. no. 4704.0.55.001. Canberra: AIHW & ABS.

AIHW GPSCU (GP Statistics and Classification Unit) 2003. General practice activity in Australia 2002–03. AIHW cat. no. GEP 14. Canberra: AIHW.

Ampon RD, Williamson M, Correll PK & Marks GB 2005. Impact of asthma on self-reported health status and quality of life: a population based study of Australians aged 18–64. Thorax 60:735–9.

Anderson RN & Rosenberg HM 1998. Age standardization of death rates: implementation of the year 200 standard. National Vital Statistics Reports; vol 47 no. 3. Hyattsville, Maryland: National Center for Health Statistics.

Baker DF, Marks GB, Walker S, Xuan W, Van der Hoek R & Hargreaves J 2003. The impact of changes to disease coding on asthma mortality and hospital morbidity statistics. Presented at Thoracic Society, Australia and New Zealand Conference, Adelaide, March.

Barros AJD & Hirakata 2003. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Medical Research Methodology 3:21.

Britt H, Angelis M & Harris E 1998. The reliability and validity of doctor-recorded morbidity data in active data collection systems. Scandinavian Journal of Primary Health Care-Supplement 16:50–5.

Britt H, Miller G, Knox S, Charles J, Valenti L, Henderson J et al. 2001. General practice activity in Australia, 2000–2001. General Practice Series no. 8. AIHW cat. no. GEP 8. Canberra: AIHW.

Britt H & Miller GC 2007. Patient-based substudies from BEACH: abstracts and research tools 1999–2006. General practice series no. 20. AIHW cat. no. GEP 20. Canberra: Australian Institute of Health and Welfare.

Dobson AJ, Kuulasmaa K, Eberle E & Scherer J 1991. Confidence intervals for weighted sums of poisson parameters. Statistics in Medicine 10:457–62.

Health Data Standards Committee 2006. National health data dictionary. Version 13. AIHW cat. no. HWI 88. Canberra: AIHW.

HIC (Health Insurance Commisssion) 2002. Medicare Benefits Schedule (MBS) group statistics reports. Canberra: HIC. Viewed 7 April 2003, <www.hic.gov.au/statistics/dyn_mbs/ forms/mbsgtab4.shtml>.

Jones A, Bentham G & Horwell C 1999. Health service accessibility and deaths from asthma. International Journal of Epidemiology 28:101–5.

McCullagh P & Nelder JA 1999. Generalized linear models. Second Edition. Boco Raton, Florida: Chapman & Hall/CRC.

NAEPP (National Asthma Education and Prevention Program) 1997. Expert panel report 2: guidelines for the diagnosis and management of asthma. Bethesda, MD: National Institutes of Health (NIH), National Heart, Lung, and Blood Institute. NIH Publication no. 97–4051.

National Centre for Classification in Health 2005. Additional diagnosis. Unpublished discussion paper submitted to the Statistical Information Management Committee, Meeting 2 August 2005, Canberra ACT, Agenda item 2.2.

NSW Health Department 2001. New South Wales health survey 1997/1998. NSW Health Department. Viewed 8 November 2002, <http://internal.health.nsw.gov.au/public-health/nswhs/methods.htm#develop>.

Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J & Powell J (eds) 1992. Cancer incidence in five continents. Vol VI. IARC Scientific Publications no. 120. Lyon: World Health Organization, International Agency for Research on Cancer.

Ringland C, Correll PK, Lim KH, Williamson M & Marks GB 2006. Hospital readmissions for asthma: a feasibility study comparing strategies for linking hospital morbidity data. Australian & New Zealand Journal of Public Health 30:435–9.

Rothman KJ & Greenland S 1998. Chapter 14. In. Modern Epidemiology. Second edition. Philadelphia: Lippincott-Raven, 243.

SAS Institute Inc. © 1999–2001. The data analysis for this paper was generated using SAS/SURVEYMEANS software, Version 8.01 of the SAS System for Windows. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. Cary, NC, USA: SAS Institute Inc.

Sears MR, Rea HH, De Boer G, Beaglehole R, Gillies AJD, Holst PE et al. 1986. Accuracy of certification of deaths due to asthma. American Journal of Epidemiology 124:1004–11.

Smyth ET, Wright SC, Evans AE, Sinnamon DG & MacMahon J 1996. Death from airways obstruction: accuracy of certification in Northern Ireland. Thorax 51:293–7.

Turrell G, Stanley L, de Looper M & Oldenburg B 2006. Health inequalities in Australia: Morbidity, health behaviours, risk factors and health service use. Health Inequalities Monitoring Series No. 2. AIHW cat. no. PHE 72. Canberra: Queensland University of Technology and the Australian Institute of Health and Welfare.

Walter SD 1978. Calculation of attributable risks from epidemiological data. International Journal of Epidemiology 7:175–82.

WHO (World Health Organization) Collaborating Centre for Drug Statistics Methodology 2006. Guidelines for Anatomical Therapeutic Chemical Classification and Defined Daily Dose Assignment. Oslo: WHO. Viewed 29 March 2006, <www.whocc.no/atcddd/>.

WICC (Classification Committee of the World Organization of Family Doctors) 1998. ICPC-2: International Classification of Primary Care. Oxford: Oxford University Press.

Wilson J 1999. An analysis of private health insurance membership in Australia, 1995. Discussion Paper no. 46. Canberra: National Centre for Social and Economic Modelling (NATSEM), University of Canberra.

Zhang J & Yu KF 2003. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. Journal of the American Medical Association 280: 1690–1.

List of Tables