Appendix 1
Supporting documentation

This page intentionally left blank

Appendix 1.1: Project resources and output

Software

The main software used in the production of this atlas were:

HealthWIZ – data analysis and mapping
Harvard Graphics – charting
Microsoft Excel for Windows – correlation analysis
Microsoft Word for Windows – word processing

Hardware

A variety of IBM compatible microcomputers were used in the production of the atlas. A HP Laser Jet 5000 Series printer was used for printing drafts of the text and maps.

Printing

The atlas was printed by Openbook Publishers, Adelaide. They were supplied with word processing documents containing the text, tables, graphs and the maps (the maps were pasted into frames in the document). The atlas was then electronically transferred to plates for offset printing, without the need for film or bromides.

Project output

Data in electronic and printed form

Separate atlases are available for each State and Territory and for Australia. For each atlas there is a companion volume comprising the data on which the maps are based: for the Australian Capital Territory, it is Volume 9.1. Both of these can be purchased from Government Info Shops in the capital cities.

The text and maps can also be downloaded for reading and printing from the Public Health Information Development Unit World Wide Web site at www.publichealth.gov.au

In addition, the text, maps and data can be accessed electronically from a CD-ROM (for Windows). On the CD-ROM, the text is in documents in Microsoft Word format. The data are in spreadsheet files in Microsoft Excel format and include all of the data mapped in the atlas, in table format as presented in Volume 9.1. Some data are also available in the HealthWIZ database.

Additional analyses will be posted to the Public Health Information Development Unit web site from time to time.

HealthWIZ software

HealthWIZ is a comprehensive health statistics database product, with a small area focus, produced by the Commonwealth Department of Health and Aged Care. It is comprised of detailed, content-rich data collections from Australia's hospital systems, cause of death registries, Medicare and social security payment systems and population censuses, together with data from administrative systems such as aged care and child care.

The data are contained on a CD-ROM and are accompanied by high performance table-building software. The menu-driven interface allows for a range of statistical calculations (age-standardised rates, confidence intervals, indices, time series data) to be undertaken to choose the most appropriate for the dataset and the needs of the user. These calculations are built into the software. The HealthWIZ software is also accessible via the World Wide Web at www.prometheus.com.au

HealthWIZ Version 4.0 comes with an integrated high performance mapping module. All the datasets and variables in the database can be mapped without the need for specialist knowledge of mapping software. All necessary digitised boundaries are included for users to be able to copy the maps to their own documents for publication.

Selected data from the atlas will be available in HealthWIZ. This includes all of the deaths and income support payments data, as well most of the hospital data, although its inclusion is subject to approval from the States and Territories. Its inclusion in HealthWIZ will allow greater flexibility in mapping the variables in the atlas, as well as many more variables from the same and other topics. The Census data, as well as the remaining health status data (the disability and handicap predictions, Total Fertility Rate), cannot be incorporated at this stage because of restrictions imposed on its use by the Australian Bureau of Statistics.

This page intentionally left blank

Appendix 1.2: Geographic areas mapped

Introduction

The following notes are intended to amplify and explain points raised in Chapter 2, *Methods* as to the areas mapped in this atlas.

Areas

Background

The basic geographic area mapped is the Statistical Local Area (SLA: SLAs are described in Chapter 2). The Statistical Local Area (SLA) is mapped in **Canberra-Queanbeyan**, as are groupings of SLAs. The groupings approximate postcode areas: see 'Areas mapped in Canberra-Queanbeyan', below. Maps have been produced in the HealthWIZ software using an approximation to Lambert's Conformal Conic Projection.

The SLA was chosen as the unit to be mapped because some datasets were only available by SLA and others were only available by postcode. It is possible to estimate data for SLAs from postcode datasets for much of Australia (basically where SLAs are larger than postcode areas, which is generally the case for areas other than in **Brisbane, Gold Coast-Tweed Heads, Townsville-Thuringowa**, **Darwin** and **Canberra**). Further, although many SLAs outside of the capital cities are of limited value for analysis (because of their large size and the often variable composition of their population) postcodes present a number of additional problems. For example, for many people living outside of a town their postcode, as used in administrative records, is the postcode of the town (i.e. the postcode of their postal address), rather than the postcode of the place in which they live. In addition, postcode areas in the country frequently cover large areas, which may not be contiguous. For example, a postcode may cover a town and the population living in a number of other towns and rural areas along a major highway, some as far as 100 or more kilometres away. Intervening towns may have a different postcode.

Areas in Canberra-Queanbeyan

In **Canberra-Queanbeyan**, SLAs are based on suburbs and are relatively small (and much smaller on average than SLAs in most other large cities). Small SLAs are likely to have smaller numbers of cases (whether of population, hospital admissions or of deaths) and these are likely to produce results (percentages, ratios) which are less reliable than those for larger areas. Throughout the atlas, estimates with small numbers of cases have not been mapped. To ensure that the majority of areas in these major urban centres are of sufficient size to produce useful results, many of the SLAs have been grouped to form larger areas. The groupings approximate (and are frequently the same as) individual postcode areas. **Table A3** shows the way in which the SLAs have been grouped.

The areas mapped for **Canberra-Queanbeyan** are shown in **Map A2** and listed in the accompanying tables. Copies of the boundaries to use as overlays with the maps in this volume are in a pocket inside the back cover.

Areas with fewer than 100 people were not mapped in any chapter (see **Table A1**). In addition, small numbers of cases were also excluded from the analysis in other chapters. For example, where the number of deaths in any area that was expected from the Australian rates was below five, the data was not mapped. Similar exclusions applied to the other data in Chapter 5 and to the data mapped in Chapter 6. The particular exclusions are noted in each chapter.

Table A1: SLAs not mapped: Population less than 100

1991-1994	1996
Belconnen-SSD Balance	Belconnen-SSD Balance
Fyshwick	Fyshwick
Hume	Gungahlin-Hall-SSD Balance
Jerrabomberra	Hume
Kowen	Jerrabomberra
Mitchell	Kowen
Parkes	Mitchell
Russell	Parkes
Weston Creek--SSD Balance	Russell
	Stromlo
	Tuggeranong-SSD Balance
	Weston Creek-Stromlo-SSD Balance

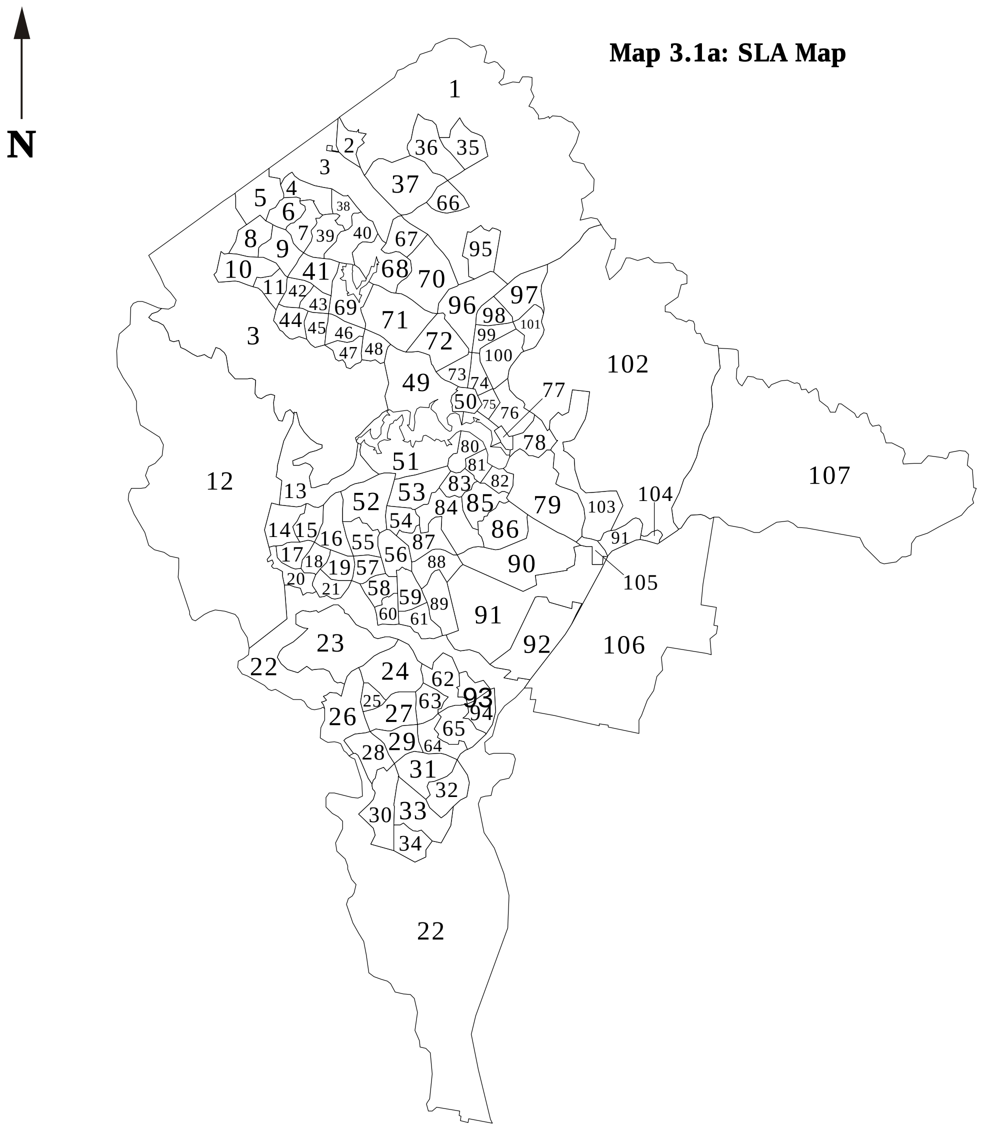
Source: Compiled from 1996 ABS Census data

Boundary changes

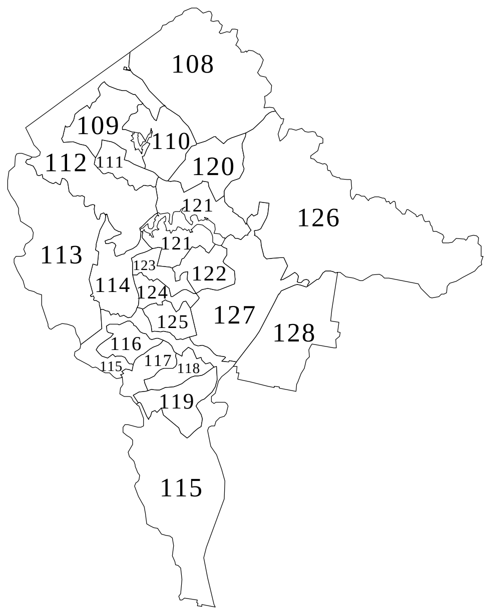
The correlation analysis reported in Chapters 5 and 6 relies on each of the datasets being collected and coded for similar spatial areas. This was not always the case in **Canberra**. As noted above, the boundaries of SLAs have changed over the period for which the datasets analysed in this atlas were collected and coded. For example, boundary changes to the SLAs of Banks, Conder and Tuggeranong SSD Balance in 1992 meant that, to be comparable, data for deaths needed to be analysed for the combined area of Banks/Conder/Tuggeranong SSD Balance. A list of the areas grouped and the name assigned to each is included in the beginning of the relevant chapter.

Map A1
Key to areas mapped for Canberra-Queanbeyan, 1996[1]

(also included as a clear film overlay inside back cover flap)



Map 3.1a: SLA Map



Map 3.1b: Postcode Map

[1]See footnotes to Table A2 (page 233) for details of differences in boundaries for areas prior to 1996

Table A2: Key to Statistical Local Areas in Canberra-Queanbeyan, 1996

Statistical Local Area Name	Area number	SLA code	Statistical Local Area Name	Area number	SLA code
Acton	49	89	Isabella Plains	29	4509
Ainslie	100	189	Jerrabomberra	91	4589
Amaroo[1]	35	239	Kaleen	70	4779
Aranda	48	279	Kambah	23	4869
Banks[2]	34	339	Kingston	82	4959
Barton	81	369	Kowen	107	5049
Belconnen Town Centre	69	459	Latham	9	5139
Belconnen-SSD Balance	3	549	Lyneham	96	5229
Bonython	28	609	Lyons	55	5319
Braddon	74	639	McKellar	68	5409
Bruce	71	729	Macarthur	93	5489
Calwell	31	819	Macgregor	8	5589
Campbell	76	909	Macquarie	46	5679
Chapman	20	1089	Majura	102	5769
Charnwood	6	1179	Mawson	59	5859
Chifley	57	1269	Melba	39	5949
Chisholm	65	1359	Mitchell	95	6039
City	50	1449	Monash	27	6129
Conder[2]	33	1549	Narrabundah	86	6219
Cook	47	1629	Ngunnawal[4]	36	6249
Curtin	52	1719	Nicholls[4]	37	6279
Deakin	53	1809	Oaks Estate	104	6309
Dickson	99	1889	O Connor	72	6389
Downer	98	1989	O Malley	88	6489
Duffy	14	2079	Oxley	25	6579
Dunlop	5	2139	Page	43	6669
Duntroon	78	2169	Palmerston[4]	66	6719
Evatt	40	2259	Parkes	80	6759
Fadden	62	2349	Pearce	58	6849
Farrer	61	2439	Philip	56	6939
Fisher	21	2529	Pialligo	103	7029
Florey	41	2619	Red Hill	84	7119
Flynn	7	2709	Reid	75	7209
Forrest	83	2789	Richardson	64	7289
Fraser	4	2889	Rivett	17	7389
Fyshwick	79	2979	Russell	77	7479
Garran	87	3069	Scullin	42	7569
Gilmore	94	3159	Spence	38	7659
Giralang	67	3249	Stirling	18	7749
Gordon	30	3289	Stromlo	12	7839
Gowrie	63	3339	Symonston	90	7929
Greenway	26	3379	Theodore	32	8019
Griffith	85	3429	Torrens	60	8109
Gungahlin-Hall-SSD Balance[4]	1	3529	Tuggeranong-SSD Balance[2]	22	8189
Hackett	101	3609	Turner	73	8289
Hall	2	3689	Wanniassa	24	8379
Harman	105	3789	Waramanga	19	8469
Hawker	44	3879	Watson	97	8559
Higgins	11	3969	Weetangera	45	8649
Holder	15	4059	Weston	16	8739
Holt	10	4149	Weston Creek-Stromlo-SSD Balance	13	8829
Hughes	54	4239	Yarralumla	51	8919
Hume	92	4329	Queanbeyan	106	6450
Isaacs	89	4419			

[1]For data sets prior to 1996, Amaroo was included in Gungahlin-Balance

[2]Banks, Conder and Tuggeranong-SSD Balance have been mapped as Banks/Conder /Tuggeranong-SSD Balance in Chapter 5 due to boundary changes in 1992

[3]For data sets prior to 1996, Dunlop was included in Belconnen-SSD Balance

[4]Ngunnawal, Nicholls, Palmerston and Gungahlin-Hall-SSD Balance have been mapped as Ngunnawal,/Nicholls/Palmerston/Gungahlin-Hall-SSD Balance in Chapter 5 due to boundary changes in 1993-95

Source: Compiled from project sources

Table A3: Key to Canberra-Queanbeyan SLA groupings, 1996

SLA group	Area no.	SLA	SLA group	Area no.	SLA
Canberra Central	121	Acton	**Tuggeranong South**	115	Banks
	121	Barton		115	Conder
	121	Braddon		115	Gordon
	121	Campbell		115	Tuggeranong-SSD Balance
	121	City	**Weston Creek**	114	Chapman
	121	Deakin		114	Duffy
	121	Duntroon		114	Fisher
	121	Parkes		114	Holder
	121	Reid		114	Rivett
	121	Russell		114	Stirling
	121	Turner		114	Waramanga
	121	Yarralumla		114	Weston
Canberra North	120	Ainslie		114	Weston Creek-Stromlo-SSD Balance
	120	Dickson	**Belconnen South**	111	Aranda
	120	Downer		111	Cook
	120	Hackett		111	Hawker
	120	Lyneham		111	Macquarie
	120	O Connor		111	Page
	120	Watson		111	Scullin
Canberra South	122	Forrest		111	Weetangera
	122	Griffith	**Belconnen West**	109	Belconnen Town Centre
	122	Kingston		109	Charnwood
	122	Narrabundah		109	Fraser
	122	Red Hill		109	Flynn
Woden North	123	Curtin		109	Fraser
	123	Garran		109	Higgins
	123	Hughes		109	Holt
Woden Central	124	Chifley		109	Latham
	124	Lyons		109	Macgregor
	124	O Malley		109	Melba
	124	Philip		109	Spence
Woden South	125	Farrer	**Gungahlin**	108	Amaroo
	125	Isaacs		108	Gungahlin-Hall-SSD Balance
	125	Mawson		108	Hall
	125	Pearce		108	Mitchell
	125	Torrens		108	Ngunnawal
Belconnen North	110	Bruce		108	Nicholls
	110	Evatt		108	Palmerston
	110	Giralang	**Kowen and Majura**	126	Kowen
	110	Kaleen		126	Majura
	110	McKellar	**Belconnen-SSD Balance**	112	Belconnen-SSD Balance
Kambah	116	Kambah		112	Dunlop
Tuggeranong North West	117	Greenway	**Stromlo**	113	Stromlo
	117	Oxley	**Eastern Fringe**	127	Fyshwick
	117	Wanniassa		127	Harman
Tuggeranong North East	118	Fadden		127	Hume
	118	Gowrie		127	Jerrabomberra
	118	Macarthur		127	Oaks Estate
	118	Monash		127	Pialligo
Tuggeranong South East	119	Bonython		127	Symonston
	119	Calwell			
	119	Chisholm	**Queanbeyan**	128	Queanbeyan
	119	Gilmore			
	119	Isabella Plains			
	119	Richardson			
	119	Theodore			

Source: Compiled from project sources

Appendix 1.3: Analysis and presentation of data

Data ranges settings

The selection of data ranges for the maps in this atlas took into account a variety of factors. These factors were:

- the data ranges used for other maps, particularly closely related maps;
- the number of areas in each range; and
- the 'balance' of the visual impact of the map.

Indirect standardisation

In comparing populations, for example the mortality of two populations, crude rates (eg. the number of deaths per 1,000 persons) may be misleading. Mortality, for example, depends strongly on age and sex. If the two areas have different age structures this variation alone may explain a difference in crude rates. The technique of standardisation is used to prevent variations in population structure from distorting differentials in events.

Indirect standardisation, used in this analysis, calculates the number of events (eg. services by GPs) which would theoretically occur if the rates for each age/sex group in a given population (the standard – in this case the population of the Australian Capital Territory) were applied to the population of interest. The result is termed the 'expected' number of events. If the actual number of events is then divided by this expected number and expressed as a percentage, we obtain the standardised ratio, a figure which is independent of population age and sex structure.

Thus the standardised ratio for a particular area will show the percentage by which it differs from the experience found in the whole population. Taking an example, the Standardised Death Ratio for deaths of males in the suburb of Ainslie was 259: that is, there were more than two and a half times the number of deaths of male residents of Ainslie aged from 15 to 64 years (159 per cent more) than would have been the case had the Australian Capital Territory (ACT) rates applied in Ainslie. In other words, the ratio was substantially above the ACT average.

The data for persons (ie. the total of females and males) has been standardised for both age and sex. That is, standardised ratios have been produced using separate details of the number of males and females in each age group. This eliminates distortion of the data which may occur where the illness or death experience of males and females is different (eg. as in the case for circulatory system disease among the population under 65 years of age). The ages used for all but the deaths data were generally each five year age group from 0 to 4 years to 80 to 84 years, and 85 years and over. For the deaths data, the ages were the five year age groups for the population aged from 15 to 64 years for all but accidents, poisonings and violence (where a separate analysis was undertaken for 15 to 24 year olds) and infant deaths. In the case of infant deaths (deaths of children under 12 months of age), the Infant Death Rate was calculated; the Infant Death Rate is the number infant deaths per 1,000 live births. Standardised ratios (SRs) were not calculated for areas where fewer than five events(deaths, admissions, etc.) were expected from the ACT rates, because of the doubtful reliability for such small numbers. All cases were, however, retained in the analysis for the preparation of capital city and State/Territory totals and ratios.

In some areas, however, high ratios are due to the relatively high proportion of Aboriginal and/or Torres Strait Islander people. This occurs because, in the methodology used, a standard population with a fixed age structure is introduced. The mortality or morbidity, etc., for a particular population (eg. people in an SLA) is then adjusted to allow for discrepancies in age structure between the standard and the particular population. When the particular population includes a sub group with a substantially different age structure and health experience (for example, mortality experience) the process is distorted. Indigenous people represent such a population. They have a substantially lower life expectancy than the total population, are a much younger population, have higher age-specific death rates at all ages and their average age at death is lower. However, since data relating to Indigenous people is not adequately identified in, for example, death or hospital statistics, they cannot be analysed as a discrete group.

The high SRs for some data for areas with a relatively large proportion of Indigenous people therefore reflect, in part, that the data has not been effectively standardised. This does not invalidate the data for these areas – on the contrary, it highlights the inequity evident in the health of Indigenous people, and the urgent need to address this inequity, as well as the need to identify Indigenous people more accurately in the statistics.

It should be noted that SRs derived for each area by this indirect method are comparable only by relation to the standard population (the State) and not directly with each other.

For variables presented as SRs the text and tables include details of whether the ratios were statistically significant ie. that they differed significantly from the standard. Whether an SR for an area differs significantly from the standard depends not only on the size of the ratio but also on the population size of the area and the overall rate for the particular event (eg. a cause of death, use of a general medical practitioner), both of which contribute to the 'expected' number of cases in an area. The same SR value in two areas which differ greatly in population size may be significantly different from the standard in the area with the larger population, but not so in the area with the smaller population.

Data sources

Table A4 shows data sources in addition to those noted in the footnotes to the tables in the earlier chapters. Further details of the HealthWIZ software (referenced in the table) are on page 229.

Table A4: Data sources

Chapter	Data sources
Chapter 4	
Tables	
4.2 to 4.6	Data for **1989** from *A Social Health Atlas of Australia 1992* . Data for **1996** is at 30 June and was compiled in HealthWIZ from data supplied by the DFACS (for all variables), DVA (Service Pension (Age) and Service Pension (Permanently Incapacitated)) and ATSIC (Community Development Employment Program data, at 30 June 1998).
Maps	As for Tables, above
Chapter 5	
Tables	
5.4 to 5.5	Compiled in HealthWIZ from data supplied by the ABS.
5.6	Data for **1988** from *A Social Health Atlas of Australia 1992* . Data for **1993** was compiled in HealthWIZ from data supplied by the ABS.
5.8 to 5.18	Data for **1985** to **1989** from *A Social Health Atlas of Australia 1992* . Data for **1992** to **1995** was compiled in HealthWIZ from data supplied by the Registrars of Deaths.
5.19	Compiled in HealthWIZ from data supplied by the ABS.
Figures	
5.3 to 5.7, 5.10	See note for Tables, above
Maps	As for Tables, above
Chapter 6	
Tables	
6.3, 6.5	With the exception of data for Queensland, data was compiled in HealthWIZ from data supplied by the AIHW from the National Hospital Morbidity Database: this database comprises data supplied to the AIHW by the State and Territory health authorities. Data for SLAs in Queensland were not available from the AIHW database and were obtained directly from the Queensland Health Department. The data was supplemented with details of the postcode or SLA of patients admitted to hospital in a State/Territory other than the State/Territory of their usual residence: these details were obtained from the individual State/Territory health authorities.
6.4	Data for **1989** (1989/90 for New South Wales) is from *A Social Health Atlas of Australia 1992* . With the exception of the data for same day patients which was from *NSW Inpatient Statistics Data Book 1989-90* for NSW and for South Australia was supplied by the Department of Human Services. Data for **1995/96**: see notes re Table 6.3, above, other than for data for same day patients which was supplied by the NSW Health Department and the South Australian Department of Human Services.
6.6, 6.9 to 6.10, 6.12 to 6.15, 6.18 to 6.23	Data for **1989** is from *A Social Health Atlas of Australia 1992* . Data for **1995/96**: see notes re Table 6.3, above.
6.7 to 6.8, 6.11, 6.16 to 6.17, 6.26 to 6.34	Data for **1995/96**: see notes re Table 6.3, above.
6.36 and 6.37	Data for **1989** from *A Social Health Atlas of Australia 1992* . Data for **1996** was compiled in HealthWIZ from Medicare statistics supplied by DHAC.
6.38	Data was compiled in HealthWIZ from immunisation rates supplied from the Australian Childhood Immunisation Register by the National Centre for Immunisation Research and Surveillance of Vaccine at the New Children's Hospital, Westmead, New South Wales.
Figures	
6.1 to 6.10	See note for Table 6.3, above
Maps	As for Tables, above
Chapter 7	
Tables	
7.3	Data for **1990/91** from *A Social Health Atlas of Australia 1992* . Data for **1996/97** was compiled in HealthWIZ from Medicare statistics supplied by DHAC.
7.4 to 7.5	Data for **1989** from *A Social Health Atlas of Australia 1992* . Data for **1995/96** (public acute hospitals) and **1997** (private hospitals) was compiled in HealthWIZ from data supplied by DHAC.
7.2 and 7.6 to 7.7	Data for **1992** from *A Social Health Atlas of Australia 1992* . Data for **1997** was compiled in HealthWIZ from data supplied by DHAC.
Maps	As for Tables, above

Note: Details of abbreviations used in the table are ABS, Australian Bureau of Statistics; ATSIC, Aboriginal and Torres Strait Islander Commission; DFACS, Department of Family and Community Services; DHAC, Department of Health and Aged Care; DVA, Department of Veterans' Affairs.

Appendix 1.4: Classification of deaths, admissions and procedures

Codes used

Causes of death are classified by the Australian Bureau of Statistics to the Ninth (1975) Revision of the World Health Organisation's International Classification of Diseases (ICD-9) which was adopted for world-wide use from 1979. The codes used for the variables mapped in Chapter 5 are listed in **Table A5**.

Diagnoses and procedures mapped in Chapter 6 are classified according to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM October 1988 Revision). External causes are classified according to ICD-9-CM Supplementary Classification of External Causes of Injury and Poisoning ('E' codes) classification codes. The codes used for the variables mapped in Chapter 6 are listed in **Table A6** and **A7**.

Table A5: ICD-9 Codes for causes of death mapped in Chapter 5

Cause of death	ICD code
All cancers [malignant neoplasms]	140-208
Lung cancer	162
Circulatory system diseases	390-459
Respiratory system diseases	460-519
Accidents, poisonings and violence	E800-E999

Table A6: ICD-9 Codes for diagnoses/external causes mapped in Chapter 6

Diagnoses /External cause	ICD code
Infectious and parasitic diseases	001-139
Cancers [malignant neoplasms]	140-208
Lung	162
Female breast	174
Psychiatric conditions	290-319
Psychoses	290-299
Neurotic, personality and other disorders	300-316
Circulatory system diseases	390-459
Ischaemic heart disease	410-414
Respiratory system diseases	460-519
Bronchitis, emphysema, asthma	490-493
Accidents, poisonings and violence	E800-E999

Table A7: ICPM Codes for surgical procedures mapped in Chapter 6

Principal procedures	Codes
All procedures	010-169; 180-695; 704-789; 792-793; 795-796; 798-869
Tonsillectomy and/or adenoidectomy	28.2, 28.3
Myringotomy [limited to 0-9 year olds]	20.01
Hysterectomy [limited females aged 30 years and over]	68.3-68.7
Caesarean section [limited to females aged 15 to 44 years]	74.0, 74.1, 74.2, 74.4; 74.99
Hip replacement	81.51, 81.53
Lens insertion	13.1, 13.2, 13.3, 13.4, 13.5, 13.6, 13.7
Endoscopies	42.23, 42.24, 44.13, 44.14, 45.13, 45.14, 45.16, 45.23-45.25

This page intentionally left blank

Appendix 1.5: Synthetic estimates for small areas

Staff of the Adelaide office of the Australian Bureau of Statistics (ABS) produced the synthetic predictions discussed and mapped in Chapter 5 as a consultancy for the Public Health Information Development Unit. The following paper prepared by the ABS describes the techniques used in production of the estimates.

Introduction

Statistics for small geographic regions are generally available only through administrative sources or the population census. Although household surveys contain much data of value, they provide estimates at a broad geographic level, usually the State or Territory level or, for some of the more populous States, for large regions. Estimates are rarely available for small areas such as the Statistical Local Area (SLA) mapped in this atlas.

Estimates produced from sample surveys are subject to two types of error: non-sampling errors which arise from errors in collecting, recording and processing the data; and sampling errors which arise because a sample, rather than the entire population, is surveyed. The sampling error tends to increase as the sample size decreases. Thus estimates produced from small samples can be subject to such high sample errors as to make them too unreliable for most practical purposes. Since household surveys typically have a small sample from large regions, it is not possible to provide direct survey estimates of suitable reliability for small regions.

Through the use of synthetic estimation techniques it is possible to produce reliable region level statistics (Marker 1999). The method of synthetic estimation was applied in predicting, at the SLA level, two characteristics from the 1995 National Health Survey (NHS):

- the number of people who had a self-assessed poor or fair health status; and
- the Physical Component Summary score from the SF-36 component of the NHS (see page 72 for details of this measure).

Predictions are also provided in this atlas of the number of people with a handicap; these estimates were produced by the ABS using a similar technique as part of another project. This technical note concentrates on the prediction of the former characteristics.

Background

Synthetic estimation predicts a value for a small geographic region based on modelled survey data and known characteristics of the region. A synthetic prediction can be interpreted as the expected value, for the variable of interest, for a 'typical' area with those characteristics. The SLA was the regional level of interest for this project (in the Australian Capital Territory and, in some cases in Queensland and the Northern Territory, SLAs were grouped; details of these groupings are contained in the relevant State and Territory atlases).

The model used for predicting small region data is determined by analysing data at a higher geographic level, in this case Australia. The relationship observed at the higher level between the characteristic of interest and predictor variables is assumed to also hold at the lower level. The predictions are made by applying the model to the small region counts of the predictors. This modelling technique can be considered as a sophisticated pro-rating of Australian level characteristic of interest across the regions in accordance with the joint distributions across the regions of the predictors.

The process of producing the predictions consists of four parts:

- preparation of data;
- model fitting;
- synthetic prediction; and
- assessing the prediction.

Data

As noted above, the two characteristics predicted were self-assessed health status and the Physical Component Summary score, both from the 1995 NHS. Self-assessed health status is provided by respondents to the survey indicating their assessment of the health status, on a scale of 'Excellent', 'Very Good', 'Good', 'Fair' or 'Poor'. The variables of interest here were those of people reporting their health as being 'Fair' or 'Poor'. The Physical Component Summary score is calculated from responses to the SF-36 component of the NHS. It is derived from a subset of items that ask respondents to the NHS aged 18 years and over, about their general physical health and wellbeing. A higher score indicates a better state of physical health and wellbeing.

Predictor data must satisfy the following criteria. It must be

- well related to the characteristic of interest;
- available from the NHS;
- available for similar time periods, both date and length of period; and be
- available at a similar geographic level, both Australia and SLA.

Sources of predictor data utilised were:

- the 1995 NHS;
- the 1996 Census of Population and Housing;
- administrative data from the Department of Family and Community Services;
- hospital separations data; and
- unreferred attendances with general medical practitioners (GPs).

One of the most important data related tasks was to identify predictors from these potential sources which satisfy the above criteria. Data considered included variables such as:

- age;
- sex;
- employment status;
- currently a student;
- income;
- receiving a Disability Support Pension;
- receiving Sickness Allowance;
- receiving the Age Pension;
- Socio-Economic Indexes for Areas derived from the Census;

- whether an inpatient at a hospital; and
- whether consulted with a GP in the two weeks prior to interview.

Many of the available variables common with the NHS differed by definition, collection methodology, reference period and geography. In such instances, appropriate adjustments were made using information obtained by comparing counts, proportions and distributions of the common variables. For example, the income variable was available to the nearest dollar from the NHS, but was available from the Census by income range only. This required the NHS income data to be classified to similar ranges. A comparison of the counts and distributions of persons across the income ranges indicated that income data from the NHS and Census were closely aligned and for the purposes of prediction could be considered well aligned. Several variables also required conversion of their geography from postcode to SLA using the 1994 Australian Standard Geographical Classification (ABS 1994).

There was, however, a fair degree of commonality in the datasets, with the NHS taken over the 1995 year, the hospital inpatient data being for 1995-96, pensioner and beneficiary data being at 30 June 1996 and the Population Census at 4 August 1996.

Model fitting

Once data preparation was completed the relationship between the characteristic of interest and the predictor variables was modelled using data from the NHS at the Australian level. The self-assessed health status and Physical Component Summary score were modelled independently.

The model applied took the linear form:

$$Y = p_o + p_1X_1 + p_2X_2 + p_3X_3 + ... + p_jX_j$$

where
Y is the characteristic of interest

X_i are the predictor variables

p_i are the coefficients which are produced from the modelling process.

In the case of the variable for self-assessed health status, the Y takes the value 1 if the individual's status was fair or poor and 0 otherwise. For the Physical Component Summary score, Y ranges in value from around 45 to 55.

The X_i predictors take the value 1 if the individual has the predictor characteristic (eg. has visited a GP in last two weeks) or 0 otherwise.

The coefficients, p_i, were estimated using the linear regression technique. An original subset of data items from the NHS were compiled that satisfied the specified criteria. The NHS data file, with the subset of data items, was randomly split into two halves with a regression model fitted to both data sets. Data items that were not important in predicting the variable of interest in either, or both, of the two models were removed. This process continued until a final linear model was obtained whereby all variables were significant ($p<0.05$) in the estimation of the response variable (characteristic of interest). Fitting the model to the split data produces a more robust final model as it reduces the probability of including a variable with high variability.

The final form of the model was then fitted to the full data set to produce regression coefficients and diagnostics which were examined using Cook's D statistic (Cook 1979) to identify any individual respondent who had undue influence on the final parameter estimates. Any 'outliers' identified were removed from the data and the model refitted.

Below is a list of variables that were included in the final models.

Self-assessed health status:
- State/Territory of usual residence;
- age (in 10 year age groups);
- sex;
- employed;
- employed (aged 18 to 24 years);
- employed (aged 25 to 34 years);
- admitted to hospital for at least one night in the last two weeks;
- consulted a general medical practitioner in the last two weeks;
- receives Disability Support Pension;
- receives Disability Support Pension (aged 18 to 24 years);
- receives Sickness Allowance;
- receives Age Pension;
- SEIFA Index of Relative Socio-Economic Disadvantage.

Physical Component Summary score:
- State/Territory of usual residence;
- age (in 10 year age groups);
- income (gross personal annual income);
- studying (currently studying full or part-time at college, university, etc.);
- employed;
- admitted to hospital for at least one night in the last two weeks;
- consulted a general medical practitioner in the last two weeks;
- receives Disability Support Pension;
- receives Disability Support Pension (aged 18 to 24 years);
- receives Sickness Allowance;
- receives Age Pension;
- SEIFA Index of Relative Socio-Economic Disadvantage.

Synthetic prediction

The prediction for an SLA was derived from the linear combination, specified by the regression coefficients, of the counts of individuals within the SLA with the predictor characteristics.

Note that for the Physical Component Summary score the predicted value for the SLA was scaled to a person level score by dividing the prediction by the number of people aged 18 years and over. The final prediction can therefore be considered as a mean score for people living in the SLA.

The predictions of poor or fair health status give an indication of the number of persons aged 18 years and over who would assess their health as poor or fair.

The predictions were age-sex standardised to remove variations between SLAs solely related to variations in age and sex.

Assessing the predictions

The models were assessed in terms of how well they predicted for individuals, SLA and larger regions (Statistical Divisions and Sub-Divisions). This involved comparing predicted values against values determined directly from the NHS. For individuals, this was the reported value, while for SLA and larger regions it was the direct survey estimate. The comparisons were made by examining plots of the predictions against the NHS reported values and estimates. The plots were checked to ensure that there was a reasonable relationship between the predictions and NHS results.

The 95% confidence intervals were calculated for the direct survey estimates and compared to the predictions. If the majority of predictions fall within the confidence intervals then there is a high level of confidence that the predictions are reliable.

This page left intentionally blank

Appendix 1.6: Additional details of cluster analysis

Introduction

Some of the descriptions of the cluster analyses were more lengthy and technical than others. Where they were considered to be too detailed and/or technical, a shortened version is shown in Chapter 8 and the full version is shown below. Those included are the health status, health service utilisation and social health status clusters for **Canberra-Queanbeyan**.

Health status clusters

The data variables available for this analysis were the variables of premature death, disability and handicap status, the Total Fertility Rate and the two synthetically predicted estimates from the 1995 National Health Survey (the Physical Component Summary and the measure of fair/poor health).

With the exception of the infant death rate (shown as the number of deaths per 1,000 live births), all of the variables were represented by age-sex standardised ratios. Missing data values (where there were fewer than five cases for any postcode group and a standardised ratio was not calculated) were substituted by zero. Legitimate zero coded values remained as zero.

The variables for infant deaths; deaths of males and females aged between 15 to 64 years; deaths of 15 to 64 year olds from cancer, lung cancer, circulatory system diseases, respiratory system diseases and accidents, poisonings and violence; deaths of 15 to 24 year olds from accidents, poisonings and violence; and years of potential life lost were excluded from the analysis because five per cent or more of the postcode areas had no cases. Thus there were five variables to analyse 21 records. Clearly this was not quite enough data.

However, a cluster analysis of all the above variables was conducted to see if it gave a sensible solution despite the lack of data. This produced a clear two cluster solution of good quality, which was not initially accepted because it was considered uninformative.

The 21 records also did not provide quite enough information for an exploratory factor analysis, since this analysis has the same data requirements as for a cluster analysis. A factor analysis was attempted using maximum likelihood extraction and oblimin rotation, which produced a two factor solution.

A second factor analysis was run using principal components extraction and varimax rotation, which resulted in a very similar two factor solution.

Since the data could support analysis by four variables, and the variable for people with a disability was the last driver of the second factor in both of the above analyses, this variable was dropped and the cluster analysis rerun on the remaining four variables. This analysis resulted in a three cluster solution of fairly ordinary quality, although it lined up very well against the IRSD.

The factor drivers of the first factor of the factor analysis solutions (people reporting their health as fair or poor and the Physical Component Summary) were entered into a cluster analysis. This produced a three cluster solution which lined up well against the IRSD, but did not discriminate very well between the Medium and Good health status groups.

By allowing the inclusion of variables with no cases for five per cent to ten per cent of the postcode areas, not all mortality variables were excluded from the analysis. Accordingly, the variables for deaths of females aged 15 to 64 years, deaths of 15 to 64 year olds from cancer and years of potential life lost were included in the analysis. Thus there were eight variables to analyse 21 records. Clearly this was not enough data.

A cluster analysis of all the above variables was tried to see if it gave a sensible solution despite the lack of data. This produced a clean three cluster solution (with Kowen and Majura not grouped) of reasonable quality. This solution was cleaner than any of the previous three cluster solutions considered, although it did not line up as well against the IRSD as any of the previous solutions examined. It was preferred because it was more informative than the two cluster solution generated, and discriminated between clusters better than any other solution. Thus the solution was accepted (see **Table 8.4** and **Map 8.2**).

Note that the Poor Status group did have higher status than the Total Fertility Rate. This result is understandable, in that females in socioeconomically disadvantaged areas have higher Total Fertility Rates.

A check with the IRSD found that, of the bottom six postcode areas for **Canberra-Queanbeyan** (as classified by the IRSD), one was not grouped and three of the remaining five (60.0 per cent) were classified to the Poor health status group in this analysis. Further, of the top three postcode areas under the IRSD, one (33.3 per cent) was classified to the Good health status group.

Health service utilisation clusters

All but one of the variables in this data set were represented by age-sex standardised ratios: the immunisation variable is of the proportion of children fully immunised at one year of age. Missing data values (postcode groups where fewer than five hospital admissions were predicted from the Australian rates) were substituted by zero. Legitimate zero coded values remained as zero.

Problems of scale can affect the analysis as more common data items will dominate the solution. To avoid these problems, the variables were standardised and the resultant z scores were entered into the analysis.

The area of Belconnen (Balance) was excluded from the analysis due to a lack of data. The variable for admissions for hip replacement was excluded from the analysis because more than five per cent of the postcode areas had no cases. Thus there were 30 variables to analyse 20 records. Clearly this was not enough data. Alternative strategies were tried in an attempt to produce a useful solution:

A cluster analysis of all the above variables was tried to see if it gave a sensible solution despite the lack of data. This produced a two cluster solution of poor quality.

An exploratory factor analysis was run on the data using maximum likelihood extraction and oblimin rotation. The analysis failed because the correlation matrix could not be inverted.

An exploratory factor analysis was run on the data using Principal Component extraction and orthogonal rotation. The analysis produced a five factor solution. It should be noted that there was not enough data to sustain a factor analysis either.

The first drivers of the five factors (admissions for myringotomy, ischaemic heart disease, neurotic, personality or other mental disorder and infectious and parasitic diseases; and the immunisation rate) were selected for entry into a cluster analysis. This analysis produced what can only be described as a garbage solution (six clusters five of size 1).

The first four drivers of the first factor (admissions to a private hospital, admissions of males, admissions of children aged 0 to 4 years for respiratory system diseases and admissions for myringotomy) were entered into a cluster analysis. The agglomeration schedule and dendogram suggested a three or four cluster solution. Both were examined and the four cluster solution was preferred because it gave a cleaner result, performed better against the IRSD, and was more informative. The solution relies on analysis of four variables over 19 cases, and is therefore supported by the data. These are described as Very Low, Low, Medium and High health service use and are shown in **Table 8.4** and **Map 8.3**.

Note that the Low service use group did have higher use of services than the high service use group for private hospital admissions and admissions for infectious diseases.

A check with the IRSD showed that, of the bottom three postcode groups for **Canberra-Queanbeyan** as classified by the IRSD, one (33.3 per cent) was classified to the High health service use group in this analysis. Of the top two postcode groups under the IRSD, neither was classified to the Very Low health service use group.

Social health status clusters

The cluster analysis technique has also been applied to a combination of the socioeconomic status and health status data sets. The results of the cluster analysis for the combination of these data sets may be useful as a summary indicator of the 'social health' status of the population of each grouping of postcode areas.

Data considered for inclusion were the demographic variables in the final model for postcode groups in **Canberra-Queanbeyan**, used to examine socioeconomic status, and the health status variables used in the final health status model. The variables excluded from the health status model because of missing data were excluded from this model also.

There were 21 postcode areas in **Canberra-Queanbeyan** for this analysis. A cluster analysis of all the above variables was tried to see if it gave a sensible solution despite the lack of data. This produced a very clean two cluster solution of very high quality. It was not initially accepted because it was considered uninformative. Alternative strategies were tried in an attempt to produce a useful solution.

An exploratory factor analysis was run on the data using maximum likelihood extraction and oblimin rotation. The analysis produced a four factor solution. It should be noted that there was not enough data to sustain a factor analysis either.

An exploratory factor analysis was run on the data using Principal Component extraction and orthogonal (varimax) rotation. The analysis produced a similar four factor solution.

The first drivers of the four factors of the oblique factor analysis solution (low income families, early school leavers, people reporting their health as fair or poor and people with a handicap) were selected for entry into a cluster analysis. This analysis produced a four cluster solution of reasonable quality. The discrimination between clusters was disappointing, but the solution compared well to the IRSD, and was informative since it was a four cluster solution.

The first drivers of the four factors of the orthogonal factor analysis solution (female labour force participation, early school leavers, people reporting their health as fair or poor and the Total Fertility Rate) were selected for entry into a cluster analysis. This analysis produced a two cluster solution of even better quality than the original two cluster solution.

The first four drivers of the first factor of the orthogonal factor analysis (single parent families, low income families, unemployed people, female labour force participation and dwellings rented from the housing authority) were entered into a cluster analysis. This analysis produced a two cluster solution of inferior quality to any of the other solutions examined.

The first factor of the oblique factor analysis consisted of only one variable (people reporting their health as fair or poor), and was therefore not suited to any further experimentation.

These analyses produced a very clean two cluster solution, or a somewhat loose four cluster solution, which lines up well against the IRSD. The latter was preferred because it is more informative. Note that ACT Eastern Fringe could have been considered as ungrouped, but it was also the lowest cluster when ranked by the mean of input variables, or when considering the IRSD, so it was considered as Very Low rather than ungrouped. The solution relies on the analysis of four variables over 21 cases, and is therefore supported by the data. The postcode groups in each cluster are listed in **Table 8.4** and shown in **Map 8.4**.

It was found that the bottom postcode group for **Canberra-Queanbeyan** as classified by the IRSD was classified to the Very Low social health status group in this analysis. Further, of the next six bottom postcode areas, four (66.7 per cent) were in the Low social health status cluster; and of the top eight postcode areas under the IRSD, seven (87.5 per cent) were classified to the High social health status group.