8 Statistical analysis

Introduction

Two sets of analyses have been undertaken to illustrate the extent of association between areas with low socioeconomic status and poor health. Correlation coefficients have been produced to indicate interdependence between the measures of socioeconomic status, health status and use of health services. Cluster analysis has been undertaken to indicate the extent to which areas display significantly similar characteristics from among the chosen measures of socioeconomic status, health status and use of health services.

Inequalities in health have traditionally been indicated by an approximation to social class, frequently based on a categorisation of occupations. The other major indicators traditionally used have included income, education, ethnicity and employment status (which allows for the inclusion of unemployed people and those not in the labour force). Measures of socioeconomic status included in this analysis include income, education, occupation, labour force status and Aboriginality.

Correlation analysis

Description

Correlation is the degree to which one variable is statistically associated with another. The correlation coefficient is a measure of the strength of this association. When high values for one variable are matched by high values for the other (or when low values are matched by low values), then they are positively correlated. Where the interdependence is inverse (ie. high values for one are matched by low values for the other), the two variables are negatively correlated.

Methods

The Pearson product-moment correlation (r) has been used in this analysis to indicate the degree of correlation between pairs of variables. Pearson correlation coefficients range from +1 (complete positive correlation) through 0 (complete lack of correlation) to –1 (complete negative correlation). As a general rule, correlations of plus or minus 0.5 or above are considered to be of meaningful statistical significance. Correlations of plus or minus 0.71 or above are of substantial statistical significance, because this higher value represents at least 50 per cent shared variation (r^2 greater than or equal to 0.5).

Correlation coefficients were calculated by comparing the value (expressed as a percentage, or as a standardised ratio) for each variable in each SLA (or postcode) with the value of each of the other variables. Correlation coefficients are generally referred to as being, for example, 'a correlation of low income families with the *paired* variable of hospital admissions of females'. However, to promote ease of reading where many correlation coefficients are quoted in the text, the word 'paired' has been omitted. For similar reasons the symbol used to indicate a correlation coefficient (r) has been omitted.

Two measures of socioeconomic status included in the analysis in this section have not been mapped. They are families receiving an income of $52,000 or more per annum and people in occupations classified as 'managers and administrators' and 'professionals'. These two measures were included as they indicate high socioeconomic status, in contrast to most other measures, which were chosen because they indicate low socioeconomic status.

The results of the correlation analysis are shown in the following tables: coefficients of from 0.5 to 0.7 and from 0.71 to 1 (both positive and negative) are highlighted in the tables, and are referred to in the individual map commentaries, as appropriate.

When discussing the results of the correlation analysis in the text, mention is often made of 'the indicators of socioeconomic disadvantage'. This reference is to variables such as those for single parent families, the unemployed, the Indigenous population and housing authority rented dwellings. References to 'high socioeconomic status' reflect the variables for high income families, female labour force participation and managers and administrators, and professionals.

The associations discussed in the text are, in general, limited to associations between the variable under discussion and the indicators of socioeconomic status from Chapter 3. This approach is largely a response to the limited space available for comment. The extent of any association with the other variables analysed can be ascertained from an examination of the correlation matrices (**Tables 8.1** and **8.2**).

Results

Canberra-Queanbeyan

There were correlations of significance at the postcode level between the measures of socioeconomic disadvantage (see Chapter 3) and a number of the health status variables. In **Canberra-Queanbeyan**, the strongest of these were generally with the variables for people reporting their health as fair or poor (as opposed to those reporting their health as being excellent, very good, or good); the Physical Component Summary (PCS, a measure of physical health); and premature death from, in particular, the combined causes of accidents, poisonings and violence (**Table 8.2**). Similarly, strong associations were also evident in the correlation analysis with the health service use variables of admissions for psychosis, accidents, poisonings and violence, Caesarean section and hysterectomy.

This page left intentionally blank

Table 8.1: Correlation matrix for SLAs in Canberra-Queanbeyan

refer to file: ch8 correlation matrices

Table 8.1: Correlation matrix for SLAs in Canberra-Queanbeyan ...*cont*

refer to file: ch8 correlation matrices

Table 8.2: Correlation matrix for postcode groups in Canberra-Queanbeyan

refer to file: ch8 correlation matrices

Table 8.2: Correlation matrix for postcode groups in Canberra-Queanbeyan ...*cont*

refer to file: ch8 correlation matrices

Cluster analysis

Description

The intention of the cluster analysis is to produce summary measures of socioeconomic status, health status and health service use at the small area level. It is useful to have this information to assist in planning and policy development activities.

It should be noted, however, that cluster analysis is an exploratory technique and, as with all such techniques, the real test of a solution is whether it makes any sense. Decisions as to the variables to be used, or the number of clusters in a solution, all impact on the final result.

The results of the cluster analysis, therefore, represent indicative groupings of areas with broadly similar characteristics among the variables analysed in each set. They will be a useful tool for some purposes: on other occasions, however, the individual variables on which they are based may also be relevant.

Methods

Cluster analysis (using the squared Euclidean measure) was undertaken by the Ward's method. This (hierarchic) clustering method seeks to partition a set of objects (eg. SLAs or, in this case, postcode groups) into a set of non-overlapping groups so as to maximise some external criterion of 'goodness of clustering', typically the extent to which the within-cluster inter-object similarities are maximised and the between-cluster similarities minimised.

In cluster analysis, 10 records (ie. postcode groups) per variable is considered desirable, with an absolute minimum of five. Had all the datasets been used in the analysis there would have been many fewer than this. A variety of techniques was used to attempt to overcome this problem, including applying a factor analysis or undertaking an experimental fit of the full data set, and using the results to reduce the number of variables included in the final analysis.

Table 8.3 lists the variables used in the analysis. The datasets used in the cluster analysis (based on boundaries in existence from 1991 to 1997) were aggregated to a common set of boundaries (1996). Where the areas differ from the 1996 boundaries, the variations are noted in the text.

Table 8.3: Variables used in cluster analysis

Socioeconomic status	**Utilisation of health services**
% single parent families	***Hospital admissions*** (Standardised Admission Ratio)
% low income families	to public acute hospitals
% unskilled or semi-skilled workers	to private acute & private psychiatric hospitals
% unemployed	to public acute & private hospitals, admissions
% female labour force participation	total
People who left school at age 15 or earlier,	of males
or who did not attend school (Standardised Ratio)	of females
% Aboriginal & Torres Strait Islander people	for infectious diseases
% Housing authority rented dwellings	for all cancers
% Dwellings without a motor vehicle	for lung cancer
Health status	for breast cancer for women aged 40 years or more
Self-reported health status	for psychoses
Physical Component Summary score [SF-36]	for neuroses
Disability and handicap status (Standardised Ratio)	for circulatory system diseases
with a disability	for ischaemic heart disease
with a handicap	for respiratory system diseases
Deaths (Standardised Death Ratio)	for respiratory system diseases in 0 to 4 year old children
Infant deaths	for bronchitis, emphysema & asthma
Deaths	from accidents, poisonings and violence
of males aged 15-64 years, from all causes	for all surgical procedures
of females aged 15-64 years, from all causes	for all surgical procedures as same day admission
of persons aged 15-64 years	for tonsillectomy and/or adenoidectomy
from cancer	for myringotomy in children aged 0-9 years
from circulatory system diseases	for Caesarean sections in women aged 15-44 years
from respiratory system diseases	for hysterectomy in women aged 30 years and over
from accidents, poisonings & violence	for hip replacements
of persons aged 15-24 years	for lens insertion in people aged 50 years or more
from accidents, poisonings & violence	for endoscopy
Years of potential life lost as a result of deaths at ages 15-64 years	***General medical practitioner services*** (Standardised Ratio)
Total Fertility Rate	for males
	for females
	Children fully immunised at 12 months

Results

Socioeconomic status clusters in Canberra-Queanbeyan

Variables considered for inclusion were those listed in **Table 8.3** under the heading *Socioeconomic status*. The ABS Index of Relative Socio-Economic Disadvantage (IRSD) was also used in the analysis, as an independent check on the solution.

Although a number of other variables were available for analysis, previous experience (Glover 1996) has shown that the inclusion of variables regarding non-English speaking background is not beneficial to the analysis. The congregation of persons of the same ethnic group does not necessarily indicate a pocket of disadvantage. Although on average we may expect these variables to also show higher levels in disadvantaged areas, their inclusion in the cluster analyses does not assist in the search for viable and sensible solutions.

The variables relating to people born in predominantly non-English speaking countries (and their proficiency in English) were accordingly dropped from the analysis. The variables for Indigenous people and dwellings with no vehicles were also dropped from the analysis since five per cent or more of the postcode areas had no cases. This left seven variable to analyses 21 records (postcode areas).

These 21 records are not theoretically sufficient to carry out a cluster analysis with seven input variables.

However, 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 an extremely clean three cluster solution of very good quality, which was accepted without further investigation. The three clusters have been labelled as High (11 postcode areas), Medium (eight postcode areas), Low (two postcode areas) socioeconomic status clusters.

The three cluster solution is supported by a comparison with the ABS Index of Relative Socio-Economic Disadvantage (IRSD) which was also available for the specified postcode areas, but was withheld from the analysis and used as an independent check on the solution. This comparison showed that, of the bottom two postcode areas with the lowest IRSD scores in **Canberra-Queanbeyan**, both were classified to the Low socioeconomic status group in this analysis; and that nine (81.8 per cent) of the 11 postcode groups with the highest scores for the IRSD were classified to the High socioeconomic status group.

Table 8.4: Composition of clusters of postcode groups in Canberra-Queanbeyan

Postcode group	Socioeconomic status	Health status	Health service utilisation	Social health status[1]
Belconnen (Balance)	High	Good	Not grouped	Low
Belconnen North	High	Medium	Medium	Medium
Belconnen South	High	Medium	Medium	High
Belconnen West	Medium	Medium	Medium	Medium
Canberra Central	Medium	Low	Low	Low
Canberra North	Medium	Low	Low	Low
Canberra South	Medium	Low	Medium	Low
Eastern Fringe	Low	Low	Low	Very low
Gungahlin	High	Good	Low	High
Kambah	Medium	Medium	Medium	Medium
Kowen and Majura	High	Not grouped	Very low	Low
Queanbeyan(C)	Low	Medium	High	Medium
Stromlo	Medium	Good	Very low	High
Tuggeranong North East	High	Medium	Low	High
Tuggeranong North West	Medium	Medium	Low	Medium
Tuggeranong South	High	Medium	High	High
Tuggeranong South East	Medium	Medium	Medium	Medium
Weston Creek	High	Medium	Medium	High
Woden Central	High	Low	High	Low
Woden North	High	Medium	Medium	High
Woden South	High	Medium	Medium	High

[1] Social health status 'clusters were produced by a joint analysis of the socioeconomic status and health status variables.
[2]Health status and Social health status cluster allocations for Pittwater are based on the combined area of Pittwater/Warringah.

Health status clusters in Canberra-Queanbeyan

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.

Several approaches resulted in a solution that, although it did not line up as well against the IRSD as any of the previous solutions examined, was more informative than the two cluster solution generated, and discriminated between clusters better than any other solution (see Appendix 1.6 for a more detailed description). 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 relatively high 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 in Canberra-Queanbeyan

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.

A number of alternative strategies were tried in an attempt to produce a useful solution. These resulted in the acceptance of a four cluster solution because it gave a cleaner result, performed better against the IRSD, and was more informative (see Appendix 1.6 for a more detailed description). 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 in Canberra-Queanbeyan

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.

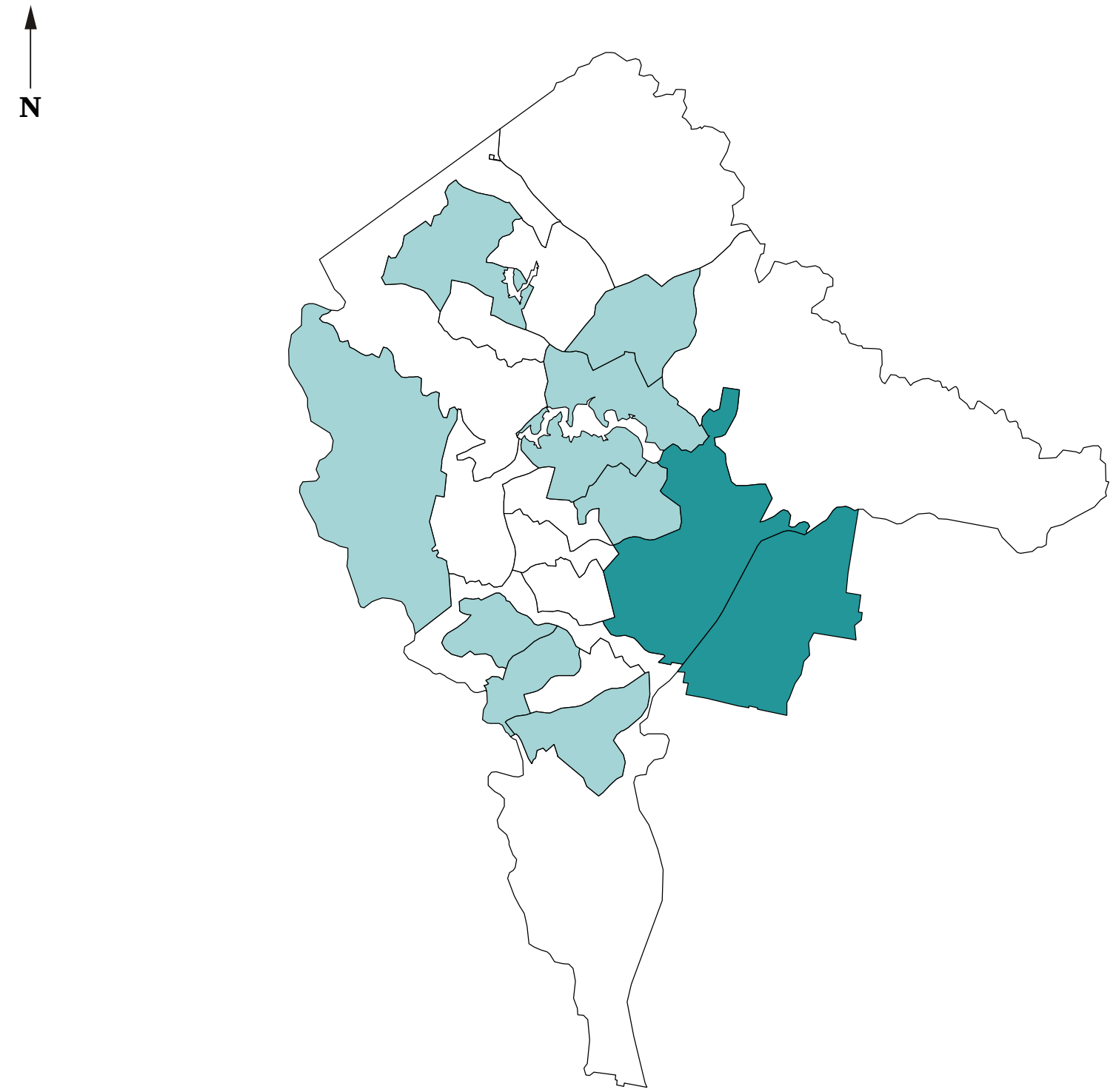
There were 21 postcode areas in **Canberra-Queanbeyan** for this analysis. A cluster analysis of all the above variables was tried to see if it gave a sensible solution despite the lack of data. This produced a very clean two cluster solution of very high quality. It was not initially accepted because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

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 (see Appendix 1.6 for a more detailed description). 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.

Map 8.1
Socioeconomic status clusters based on postcode areas, Canberra-Queanbeyan, 1994

clusters of areas with generally similar socioeconomic status characteristics



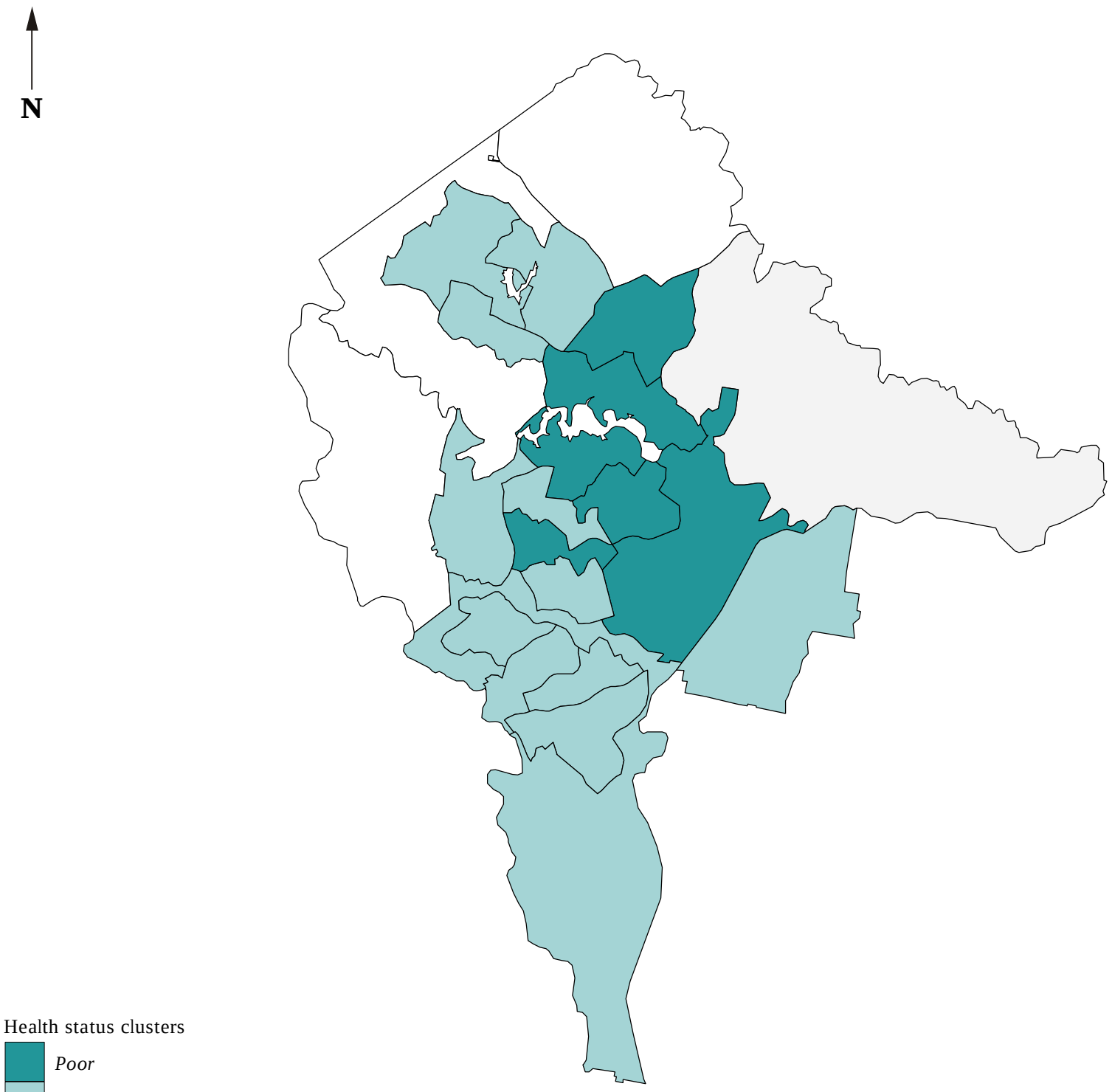
Source: See Data sources, Appendix 1.3

Details of map boundaries are in Appendix 1.2

National Social Health Atlas Project, 1999

Map 8.2
Health status clusters based on postcode areas, Canberra-Queanbeyan, 1994

clusters of areas with generally similar health status characteristics



Health status clusters

- *Poor*
- *Medium*
- *Good*
- *Not mapped*[*]

[*]*Areas not mapped include Kowen/Majura, which was not allocated in the cluster analysis*

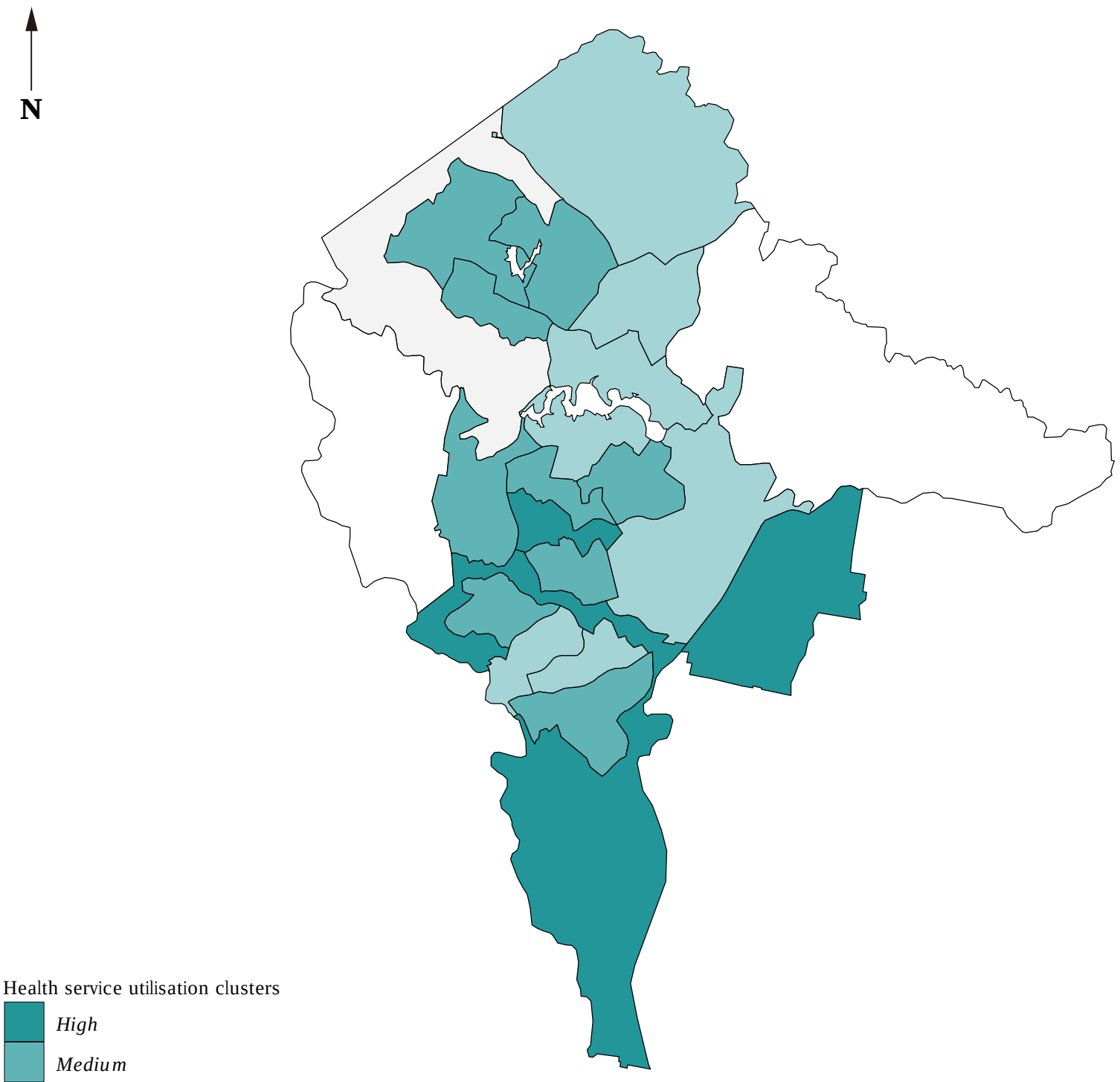
Source: See Data sources, Appendix 1.3

Details of map boundaries are in Appendix 1.2

National Social Health Atlas Project, 1999

Map 8.3
Health service utilisation clusters based on postcode areas, Canberra-Queanbeyan, 1994

clusters of areas with generally similar health service utilisation characteristics



*Areas not mapped include Stromlo, which was not allocated in the cluster analysis

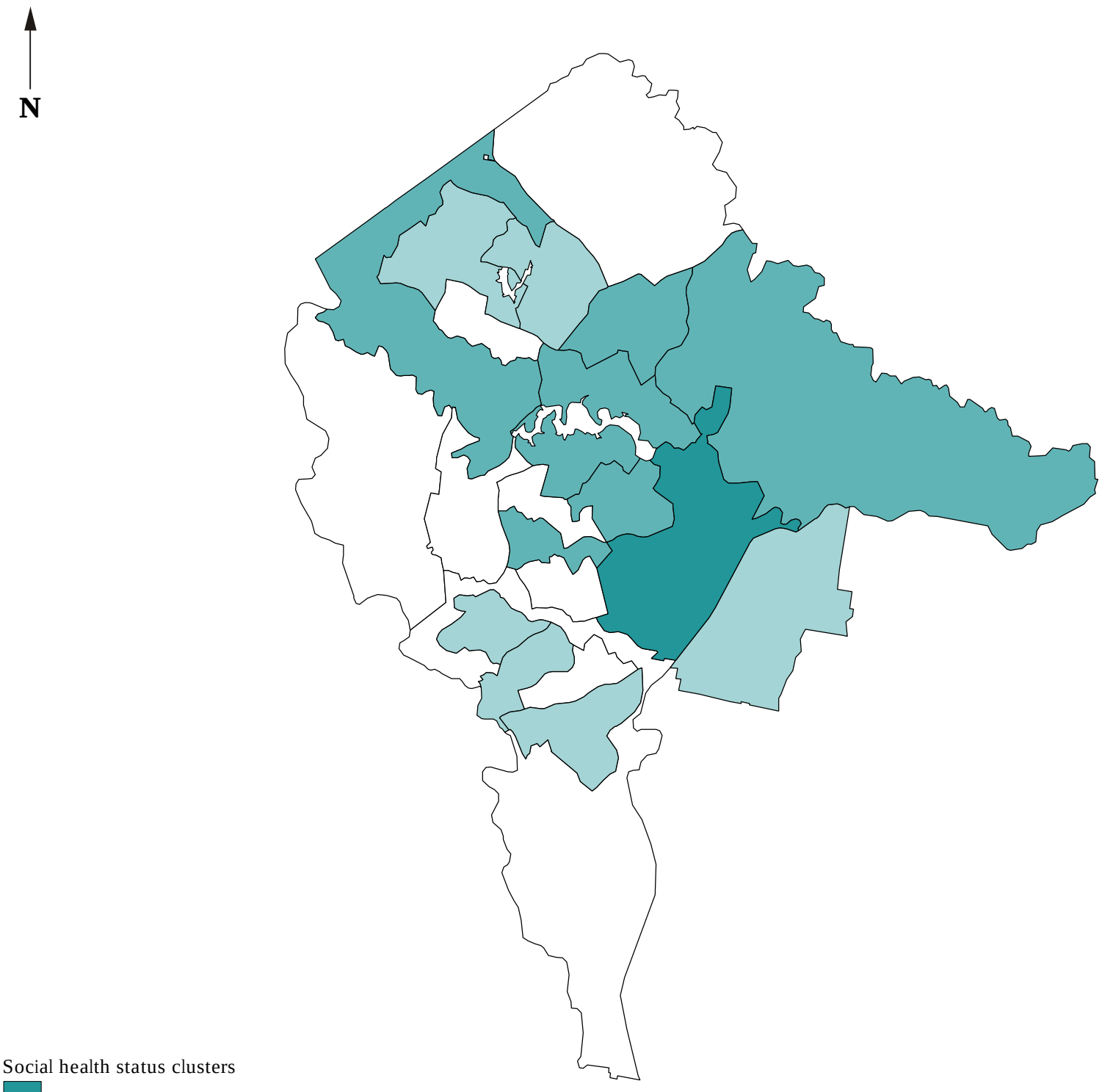
Source: See Data sources, Appendix 1.3

Details of map boundaries are in Appendix 1.2

National Social Health Atlas Project, 1999

Map 8.4
Social health status clusters based on postcode areas, Canberra-Queanbeyan, 1994

clusters of areas with generally similar social health status characteristics



Source: See Data sources, Appendix 1.3

Details of map boundaries are in Appendix 1.2

National Social Health Atlas Project, 1999

This page left intentionally blank