1351.0.55.015

Australian Bureau of Statistics

Research Paper

Socio-Economic Indexes for Areas: Introduction, Use and Future Directions

www.abs.gov.au

Research Paper

Socio-Economic Indexes for Areas: Introduction, Use and Future Directions

Pramod Adhikari

Analytical Services

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THURS 28 SEP 2006

ABS Catalogue no. 1351.0.55.015
ISBN 0 642 48252 7

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Dr Pramod Adhikari, Analytical Services Branch, on Canberra (02) 6252 7646 or email <pramod.adhikari@abs.gov.au>.

CONTENTS

SOCIO-ECONOMIC INDEXES FOR AREAS: INTRODUCTION, USE AND FUTURE DIRECTIONS

Pramod Adhikari
Analytical Services

ABSTRACT

The Socio-Economic Indexes for Areas (SEIFA) are measures which summarise a range of socio-economic variables associated with disadvantage. These indexes are compiled at the Census Collection District (CD) level, and may be used to rank CDs according to the general socio-economic well-being of residents. In this paper we discuss three important features of SEIFA. First, SEIFA scores are a measure of *relative* disadvantage. Second, SEIFA scores are *areal level* indexes and should not be presumed to apply to *all* individuals living within the area. Third, SEIFA scores are calculated at the CD level and great care is required when interpreting scores which have been aggregated to larger geographical areas. We provide examples of the use of SEIFA to analyse the distribution of relative disadvantage within larger areas. Using data from the *National Health Survey 2004–05*, we also show that SEIFA scores correlate with the proportion of people living in an area who report poor health, obesity and other health risk factors.

1. INTRODUCTION TO THE INDEXES

The first SEIFA was released in 1990 using data from the 1986 Census of Population and Housing and consisted of five indexes. Using the same methodology to derive the indexes, there have been another three SEIFA releases for the 1991, 1996 and 2001 censuses. In 1986, 1991 and 1996 five indexes were constructed. For SEIFA 2001, as a result of a review of the utility of the indexes as well as the methodology used to derive the indexes, only four indexes were constructed. Two indexes – the Urban Index of Disadvantage and the Rural Index of Disadvantage – were replaced by the Index of Relative Advantage/Disadvantage.

SEIFA indexes summarise a number of socio-economic variables that represent disadvantage in an area. This single measure can be used to rank CDs to identify areas that are more or less disadvantaged relative to others. These indexes measure different aspects of socio-economic conditions at a CD level and have been derived using a multivariate technique known as 'Principal Components Analysis'. This technique summarises the information from a variety of social and economic variables into a single measure. Different sets of CD level variables are used to construct different SEIFA indexes.

There are three factors which the indexes do not represent well. First, the indexes contain only limited information about accumulated wealth. Second, an area's infrastructure such as schools, community services, shops and transport is not represented by the indexes. Third, the indexes do not capture the difference in cost of living across different areas. The Census of Population and Housing does not collect information about these three factors, and so it is impossible to include them in the construction of the SEIFA indexes.

The four SEIFA 2001 indexes are:

- the Index of Relative Socio-Economic Disadvantage,
- the Index of Relative Socio-Economic Advantage/Disadvantage,
- the Index of Economic Resources, and
- the Index of Education and Occupation.

In the coming sections, we provide a brief description of each index.

1.1 Index of Relative Socio-Economic Disadvantage (IRSD)

The IRSD summarises Census variables, at the CD level, that are considered indicators of disadvantage. Social disadvantage is typically associated with low income, high unemployment and low levels of education. The list of variables and their corresponding weights (also known as loadings) are listed in the Appendix. Since this index only summarises variables that indicate disadvantage, a low score indicates that

an area has many low income families, people with little training and working in unskilled occupations and may be considered as disadvantaged relative to other CDs. A high score implies that the area has few families with low incomes and few people with little or no training and few people working in unskilled occupations. These areas with high index scores may be considered less disadvantaged relatively to other CDs. It is important to understand that a high score reflects lack of disadvantage rather than advantage or high advantage. To find CDs that are relatively more advantaged, the Index of Relative Socio-Economic Advantage/Disadvantage should be used.

1.2 Index of Relative Socio-Economic Advantage/Disadvantage (IRSAD)

The IRSAD summarises CD level variables that represent both advantage and disadvantage. For example, the index includes variables such as high income, low income, professional occupation, as well as people employed in unskilled occupations. Inclusion of both types of variables – indicators of advantage and disadvantage – in a single index, allows the index to be used as a measure of advantage and disadvantage in a continuum. Areas with a low index score can be categorised as relatively disadvantaged areas and areas with higher scores can be categorised as relatively advantaged areas.

1.3 Index of Economic Resources

This index summarises variables that relate to family income, rental and mortgage payments, and dwelling size. Areas that have a large proportion of families with high annual income, or households paying high rent, or paying high mortgage, dwellings with four or more bedrooms, or small proportion of families on low income, or paying low rent will have higher scores. Areas with low index scores can be considered relatively disadvantaged compared to other areas with higher index scores.

1.4 Index of Education and Occupation

This index summarises variables at the CD level in terms of the people who are unemployed, their level of qualification and if employed, the type of jobs they are employed in. High scores on this index means that the area has higher proportion of people who are well educated, or who are employed in professional occupations, or proportionately fewer unemployed people, or fewer people in low skilled jobs.

1.5 Deciding which index to use

As there are four indexes available for each CD, with each index summarising a different set of variables, researchers should carefully choose which index is suitable for their particular purpose. Of the four indexes, the most widely used indexes in

health and social research are the Index of Relative Socio-Economic Disadvantage and the Index of Relative Socio-Economic Advantage/Disadvantage. As we have explained already, if the research question involves finding areas that are advantaged, we recommend using IRSAD. With IRSAD, areas which have low scores are relatively disadvantaged and areas that have high scores are relatively advantaged. With IRSD, we can only say that low scoring areas are relatively more disadvantaged compared to other areas with high IRSD scores. Since component variables for IRSD only include measures of disadvantage, it is possible for some areas to score low in IRSD (relatively more disadvantaged) as well as high in IRSAD (relatively more advantaged).

2. NOTION OF SOCIO-ECONOMIC DISADVANTAGE

Socio-economic disadvantage is a relative concept with many, often interrelated, dimensions to it. Some would argue that access to material wealth makes a community advantaged. Others would argue that despite the access to material wealth, if the community is rife with drug problems, or has a very high crime rate then the community could be categorised as disadvantaged. Many studies have been conducted around the world in the area of social capital and how individuals interact, but unfortunately many of these community-based variables are not collected at Census time and cannot be used in the SEIFA construction. Based on international research and also the type of information ABS collects during the Census, we define disadvantage in terms of individuals' access to material and social resources, and their ability to participate in society.

Persons who have better education tend to have jobs which are higher paid. With higher education and high paying jobs comes prestige. Although the terms prestige and socio-economic status are often used interchangeably, there is a theoretical difference between the two. Krieger (2006) uses terms such as 'resource-based measure' and 'prestige-based measure' of socio-economic position. Resource-based measures refer to material and social resources and assets that include income, wealth and education. Prestige-based measures refer to individuals' status in social hierarchy and are related to prestige rather than wealth measures. However, high rewarding job/low rewarding job is also a relative concept. A job which is considered high paid in one country or in one community may not be considered equally high paid in other countries or communities. A job paying below average weekly earnings in Australia might be considered a highly paid job in many Eastern European countries where the employment rate is very high and the wages are low. The economic situation of a country may alter the concept of high paid jobs. During an economic boom, people may be able to afford to choose the type of job they want, while during recession, people might be happy just to have a job. Socio-economic status therefore is a relative concept.

Townsend (1987) summarises socio-economic deprivation as a state of disadvantage which can be observed relative to the local community or the wider society or nation which an individual, family or group belongs. The difficulty of identifying a measure of disadvantage – social, economic, health, spiritual – has led to the development of numerous socio-economic indexes. Most of these indexes include at least three main characteristics. These are: employment, education and financial well-being.

Measures of disadvantage or deprivation for small areas have been developed in many countries (see Health Information Partnership – Eastern Ontario Region, Research Report 2005 for Canada; Krieger et al., 2003 for the United States; Central Bureau of Statistics 2003 for Israel; Salmond and Crampton 2001, 2002 for New Zealand;

Department of the Environment, Transport and Regions 2000 for the United Kingdom). Many of these studies use a similar method to that used by the ABS to construct its SEIFA indexes. The value of such area level measures in explaining health differentials has been shown in many health studies (Glover, Hetzel and Tennant 2004; Krieger et al., 2003; Barnett 2000; Ross et al., 2000; Yu, Robertson and Brett 2000; Glover, Harris and Tennant 1999; Sorlie, Backlund and Keller 1995). Some authors note that such derived variables capture group properties, and these group-level and individual-level variables interact in shaping health and disease of people (Diez–Roux, 1998; Diez–Roux et al., 1997). This is a very important concept. This means that these derived variables not only measure aggregated compositional properties of the individuals at the CD level, but also the contextual properties of that area.

The contextual properties of an area may be shaped by the properties of the group. If an area has a very high supply of governmental rental housing, people from other CDs might move there. In this case, the contextual factor is shaping the properties of the population in that CD. This interrelation between the contextual property of an area and the aggregated properties of people living in an area means the interpretation of the SEIFA score is not straightforward.

The relative disadvantage (or advantage) that SEIFA indexes summarise at CD level should therefore be used as contextual variables and not as indexes for individuals living in that area. If we attempt to explain individual level disadvantage from area level measure such as SEIFA then we are assuming that the relationships observed for areas hold for individuals. As we know people living in a CD are not homogeneous – everyone living in a disadvantaged CD will not all be equally disadvantaged. There will be some people who are less disadvantaged than others even in the most disadvantaged areas. The ecological fallacy is a result of the assumption that relationships observed for areas also hold for the individuals who live there. If we assign an area level index to an individual then there is a risk of an ecological fallacy.

3. HOW THE SEIFA INDEXES ARE CALCULATED

All the indexes have been created using a multivariate technique called 'Principal Component Analysis' (PCA). PCA is a data reduction technique that involves reducing a large number of related or correlated variables into a smaller set of derived or transformed variables that account for as much of the (original) total variation as possible. These transformed variables are called 'principal components'. There can be as many principal components as there are variables.

The first principal component accounts for the largest proportion of the variation in the original data set. Successive principal components are extracted in such a way that they are uncorrelated with each other and account for successively smaller amounts of the remaining total variation. While it is possible to extract as many principal components as there are original variables, the goal in PCA is to reduce the dimensionality in the data. In other words, we want to summarise a large number of related variables to a small number of meaningful groups of variables. Depending upon the purpose of the summary measure and how the variables group together, one or more of the principal components are used to create the index. For example, the SEIFA Index of Relative Socio-Economic Disadvantage (IRSD) is obtained from the first principal component, which is a linear transformation of 20 disadvantage indicator variables. For more detail on the technical method see the ABS publication *Census of Population and Housing: Socio-Economic Indexes For Areas (SEIFA)* (ABS cat. no. 2039.0.55.001).

The variables for the models are selected based on the notion of disadvantage as discussed earlier and the availability of indicator variables in the Census. Once a list of variables is identified, then we assess the face, content and construct validity. Face validity requires that the index appears relevant to the construct we wish to measure. Face validity can be established by seeking comments from the community. In creating 2001 SEIFA, we sought views from a number of stakeholders on whether the variables that are included in the creation of SEIFA scores measures disadvantage (or advantage) at the area level. Large scale face validity was conducted by Walker and Hiller (2005) in South Australia for majority of variables included in the IRSD score. The results from the survey showed that 50% or more respondents in South Australia agreed that the variables included in the IRSD were relevant measures of disadvantage at the area level.

The theory behind content validity, as opposed to face validity, is that experts are aware of subtle qualities of the index that the general population may not be aware of. ABS sought expert advice from five academics/professionals in this area and their view was that the SEIFA scores included the relevant variables that measure area level disadvantage. Construct validity refers to the degree to which inferences can legitimately be made from the theoretical construct, which is derived using the

disadvantage indicators. Many studies have shown that SEIFA scores are related to health status (Yu, Robertson and Brett, 2000; Glover, Harris and Tennant, 1999). In the coming sections, we will test the construct validity using the National Health Survey 2004–05, to see whether more people from most disadvantaged areas possess poorer health or other health risk factors. The observable gradients – a graded change prevalence of health risk factors, as the areas become more disadvantaged – would indicate that the SEIFA scores are associated with area level disadvantage.

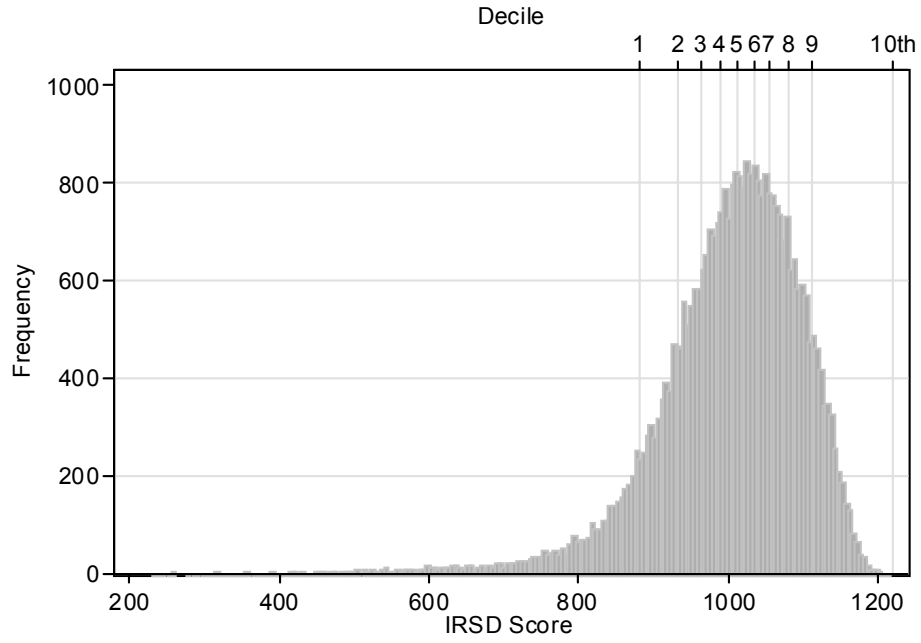
For aesthetic reasons, we standardise the principal component scores to have a mean of 1000 across all CDs and a standard deviation of 100. Even though the index scores have a mean of 1000, they should always be used as ordinal variables. An area with an index score of 500 is not necessarily twice as disadvantaged as another area with a score of 1000. We can say, however, that an area with a lower index score is relatively more disadvantaged than another area which has a higher index score. With the help of an index score, we can rank areas to see the extent of relative level of disadvantage. For example, we can say that areas with the lowest 10% of index scores are relatively more disadvantaged than those areas which have the highest 10% of scores.

4. THE DISTRIBUTION OF INDEXES ACROSS AUSTRALIA

Since CDs are the unit of analysis, each CD gets a score. Each CD will have four index scores attached to it, for the four SEIFA indexes. How the index scores are distributed across Australia is shown in the following histograms.

The histogram of IRSD scores shown in figure 4.1 by decile cut-off values shows that nearly four-fifths of IRSD scores are in the ranges from 880 to 1110. That is, the lowest 10% scores are smaller than 880 and the highest 10% scores are larger than 1110. If we look at the histogram, we notice that IRSD scores in the 4th, 5th, 6th, or 7th deciles have decile cut-off values that are very close to each other. This pattern of distribution of IRSD score means that comparing areas in terms of relative disadvantage can be best made if we compare areas between extreme IRSD deciles.

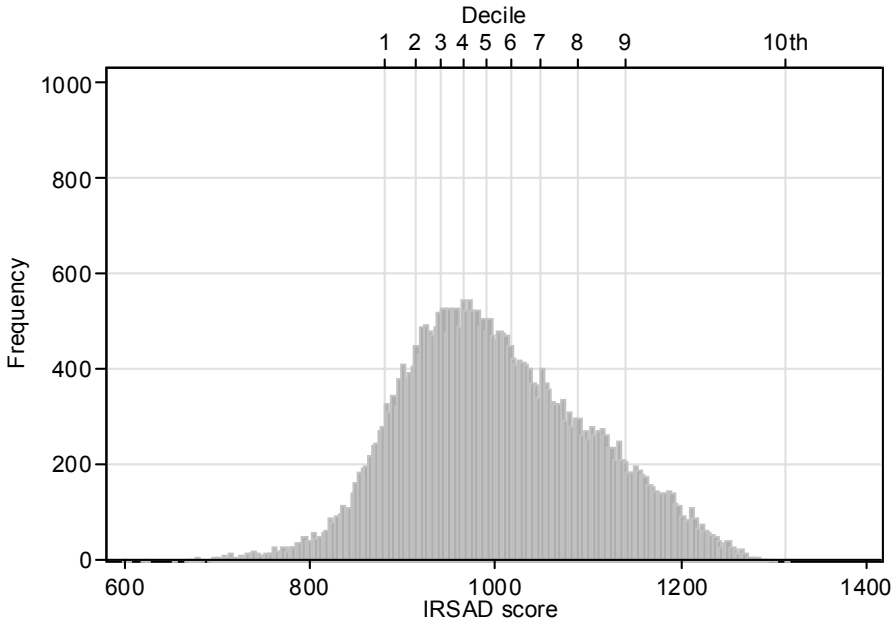
4.1 Distribution of IRSD score, with decile cut-off values



The discriminatory power of IRSD score lies at the extreme values of the IRSD scores. For example, if we are comparing an area that has a score in the first decile to another area which has a score in 10th decile, we may find these two areas are very different in their characteristics. However, if we compare two areas whose scores are in the 4th and 7th deciles, their differences may not be that obvious. Using IRSD scores, we cannot say that an area which has an index score of 600 is twice as disadvantaged as another area which has a score of 1200. However, we can say that the area with a lower score is relatively more disadvantage than the other one which has the higher IRSD score.

The histogram of IRSAD score shown in figure 4.2 is different to that one for IRSD shown in figure 1. The decile cut-off values for IRSAD scores are more dispersed than for IRSD. However, the scores that fall in 2nd to 7th deciles are closer to each other than those scores which are in the 8th or 9th deciles. Because of the wider dispersion of the scores, if users wish to compare areas according to decile of SEIFA scores, it is probably better to use IRSAD than IRSD. However, users should also note that IRSD and IRSAD use different sets variables.

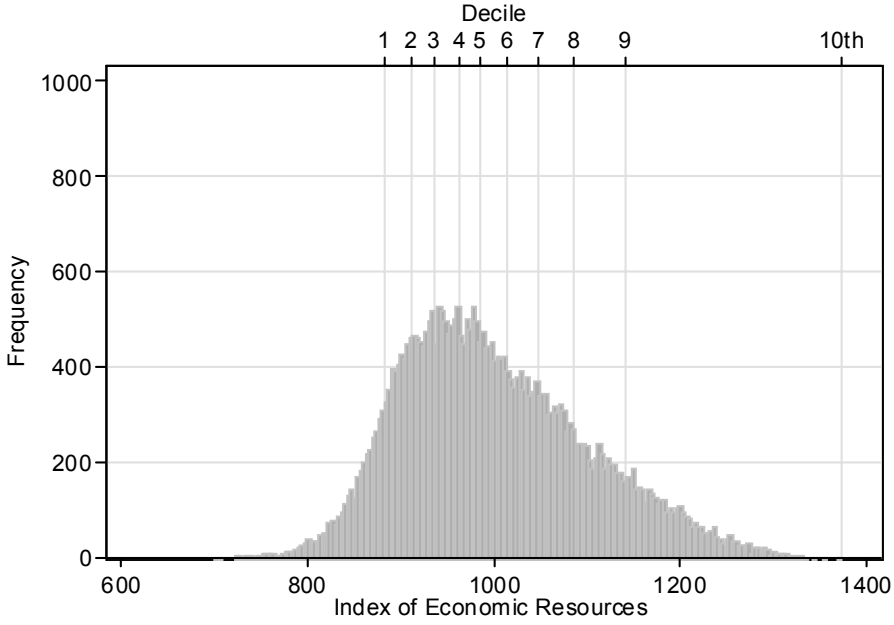
4.2 Distribution of IRSAD score, with decile cut-off values



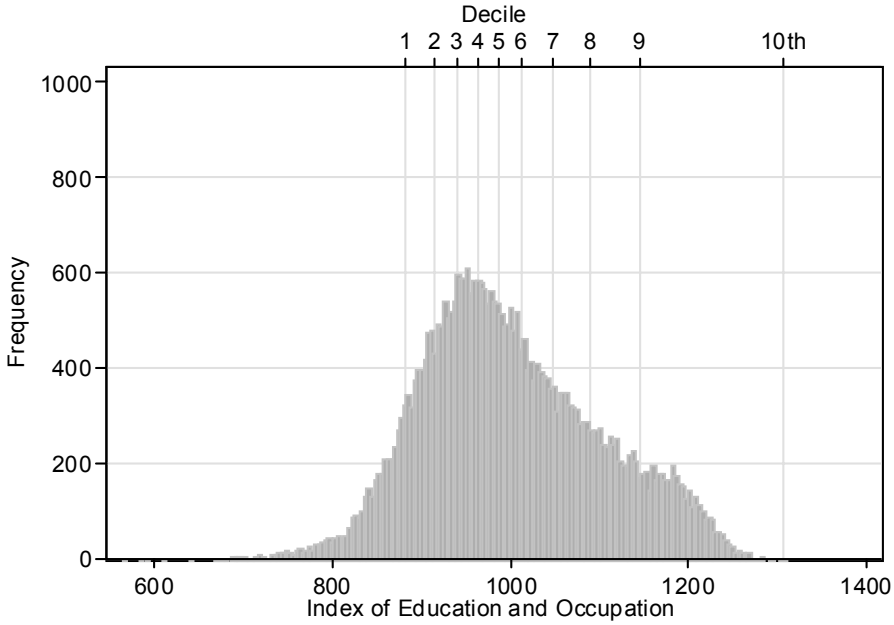
The following two graphs (figures 4.3 and 4.4) show the distribution of Index of Economic Resources and Index of Education and Occupation scores. These two index scores have similar distributional properties, with nearly identical decile cut-off values. These scores have some similarities with IRSAD distribution, with a greater spread of CDs in the higher decile (i.e. more advantaged) end of the distribution.

Since ABS produces four indexes, it is up to the researchers to decide which index is most suitable for their research project. There may be instances when use of more than one index could be of value and a particular index would be more suitable for some research than others.

4.3 Distribution of Index of Economic Resources, with decile cut-off values



4.4 Distribution of Index of Education and Occupation, with decile cut-off value



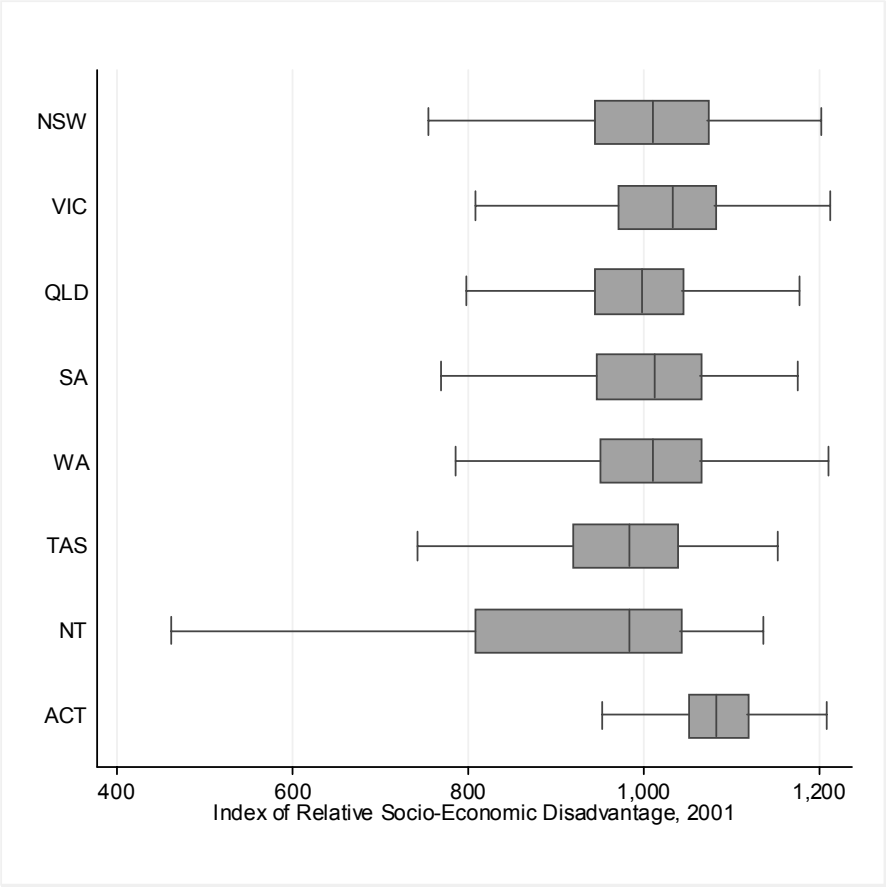
5. DISTRIBUTIONAL ANALYSIS OF SEIFA SCORES

In this section we show how we can use SEIFA index scores to compare the relative disadvantage between areas by examining the distribution of scores within a geographical region (such as state or Statistical Division).

A boxplot (also called a box and whisker plot) is ideal for visually comparing similar distributions. In the boxplot:

- The ends of the box are the upper and lower quartiles (top 25% and bottom 25% respectively). Hence, the box represents the 50% of values that lie between 25th and 75th percentiles.
- The vertical line inside the box is the median value (50% of all scores are below this value).
- The whiskers are the two lines outside the box that extend to the highest and lowest observations. The outliers, or the extreme values are not included in the whiskers.

5.1 Distribution of 2001 IRSD scores



5.1 Distribution of IRSD scores across the States and Territories

As can be seen in figure 5.1, the Australian Capital Territory has the highest median score. Furthermore, the lowest value of IRSD score for the ACT is higher compared to other states/territory, and that the IRSD scores have very little variation between the lower and upper quartiles. For the Northern Territory, however, the median IRSD score is one of the lowest, and it has many areas with low IRSD scores, with a wide range of IRSD scores.

Although boxplots show the ranges of scores, it is still not clear which state/territory has the disproportionately higher number of areas that are relatively disadvantaged (areas with the lowest 10% of IRSD scores). In table 5.2 we present the proportions of CDs for each state/territory that are in the first IRSD decile.

5.2 Distribution of CDs in the first decile by State/Territory, 2001 IRSD[(a)]

State/Territory	*Percent of CDs in the first decile*
New South Wales	10.9
Victoria	8.0
Queensland	8.9
South Australia	12.0
Western Australia	9.8
Tasmania	15.9
Northern Territory	28.3
Australian Capital Territory	1.6
Australia	10.0

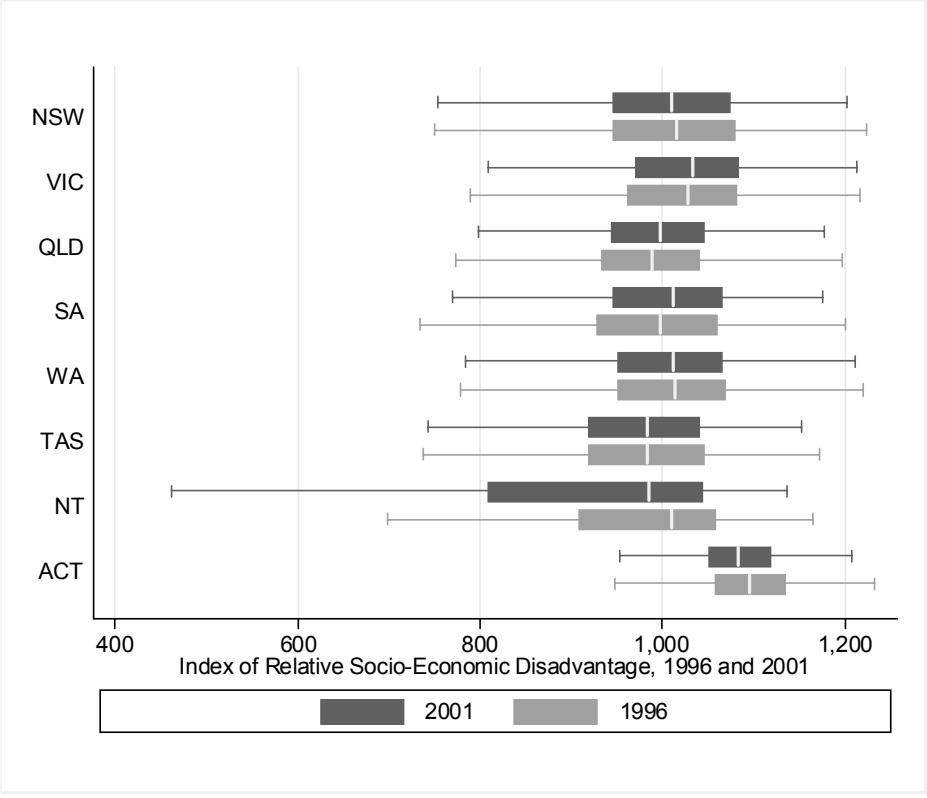
(a) The decile cut-off value is for the whole of Australia.

The data in table 5.2 reinforce the message from figure 5.1 in that, the Australian Capital Territory is the least disadvantaged area and the Northern Territory is the most disadvantaged area. Less than 2% of CDs in ACT have IRSD scores that are in the lowest decile compared to the Northern Territory, which has an over-representation of CDs with the lowest IRSD score, with nearly 30% of all its CDs in the lowest 10% of IRSD scores.

5.2 Comparison of IRSD scores across the States and Territories and Censuses

Information similar to that in figure 5.1 and table 5.2 can be produced for two Censuses if we want to compare how relative disadvantage has changed across time. We should be careful when comparing scores across time as there could be a number of reasons why they differ. These include the different definition of variables in the model, different number of variables in the model, or the area itself might have changed in terms of its population composition, or even the CD may have been divided into two. However, even if the variable definition changes or the number of variables changes across time, if we are looking at the relative disadvantage of an area, the decile ranks can be used to compare the areas, instead of the actual score. The example we have shown in figure 5.3 is for IRSD. The same variables were included in the construction of IRSD in 1996 and 2001. Comparison of these scores is less risky than comparing scores which used different sets of variables at two different time points.

5.3 Distribution of IRSD across time, 1996 and 2001



We see from figure 5.3 that compared to 1996, states such as Victoria, Queensland and South Australia had slightly higher median IRSD score in 2001. The Northern Territory and the Australian Capital Territory both have a marked decrease in the

median IRSD score in 2001. However, the median scores for these states and territories only tell part of the story.

In table 5.4, we have shown the proportion of CDs which fall in the first decile for the two Census years by state/territory. Compared to 1996, more CDs were in the lowest decile in 2001 in all jurisdictions except Queensland, South Australia and the Australian Capital Territory. It should be noted that, by design, there will be 10% of CDs across Australia with the lowest 10% of IRSD scores. This means that, if some states/territories have a smaller percent of CDs with the lowest 10% of IRSD scores, other states/territories will have to have larger proportion of CDs with the lowest 10% of IRSD scores.

5.4 Distribution of CDs in the first decile by State/Territory, 1996 2001 IRSD[(a)]

	Percentage of CDs in the first decile	
State/Territory	*1996*	*2001*
New South Wales	9.7	10.9
Victoria	7.3	8.0
Queensland	10.0	8.9
South Australia	14.6	12.0
Western Australia	9.2	9.8
Tasmania	14.4	15.9
Northern Territory	22.6	28.3
Australian Capital Territory	2.1	1.6
Australia	10.0	10.0

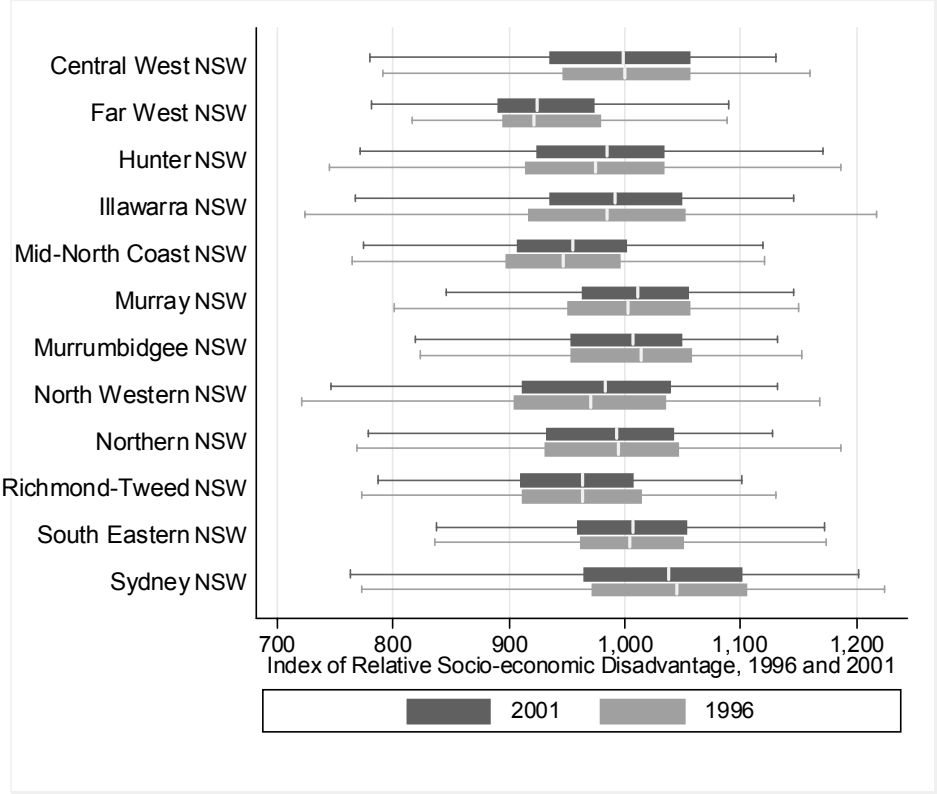
(a) The decile cut-off value is for the whole of Australia.

The strength of area-based disadvantage scores such as SEIFA lies in the distributional analysis such as those presented in figures 5.1 and 5.3 and tables 5.2 and 5.4. If we just look at the mean or median at the state/territory level, we would not be able to visualise what is happening with the areas at two different times.

5.3 Comparison of the 1996 and 2001 IRSD scores in New South Wales by Statistical Divisions

In Section 5.2 we looked at the distribution of IRSD scores by state/territory. We can perform the same kind of analysis within New South Wales at the Statistical Division (SD) level. In figure 5.5 we present the interquartile ranges, median, maximum and minimum values using boxplots for two Census periods. We notice that there is variability in the IRSD scores across SDs; and the SD of Far West New South Wales appears to be relatively the most disadvantaged SD.

5.5 Comparison of the IRSD scores 1996 and 2001 in New South Wales by Statistical Division



As explained already, one of the methods to compare the relative disadvantage of an area is to see how many of the CDs within an area are in the lowest decile of IRSD scores across time. The data in table 5.6 show that in seven of the twelve SDs in New South Wales, their proportion of CDs in the lowest decile of IRSD scores increased in 2001 compared to 1996. Statistical Divisions of Far West, Mid-North Coast and North Western were relatively most disadvantaged; with more than one-sixth of all CDs in these SDs falling in the first IRSD decile at both the Census periods.

5.6 Distribution of CDs in the first decile by Statistical Division, New South Wales, 1996 and 2001 IRSD[(a)]

	Percentage of CDs in the first decile	
State/Territory	*1996*	*2001*
Central West	10.2	12.1
Far West	18.3	19.4
Hunter	14.5	13.1
Illawarra	14.0	12.7
Mid-North Coast	18.3	16.1
Murray	4.4	5.4
Murrumbidgee	7.6	10.2
North Western	18.4	17.6
Northern	10.8	11.8
Richmond–Tweed	13.0	14.4
South Eastern	6.0	5.6
Sydney	7.5	9.7
New South Wales	9.7	10.9

(a) The decile cut-off value is for the whole of Australia.

The data also show that less than 10% of CDs in the Murray, South Eastern, and Sydney Statistical Divisions in New South Wales were in the lowest IRSD decile for both Census periods. If we define disadvantaged areas as those areas with their IRSD score in the 1st decile, then the data show that Murray, South Eastern, and Sydney SDs were relatively less disadvantaged at both censuses. On the other hand, less than 8% of CDs in the Murrumbidgee SD had IRSD scores that fell in the lowest decile in 1996, compared to little over 10% in 2001, indicating a slight increase in relative disadvantage in 2001. Just looking at the data does not tell the likely cause of this increase in disadvantage. The knowledge about the local area is very important when there is increase or decrease in the level of relative disadvantage in an area. If an area suddenly becomes relatively less disadvantaged at one point in time, one needs to be very careful in drawing conclusion as changes in SEIFA scores for a one or two CDs across time could make the percent change look higher, if the area under study has only few CDs.

6. USE OF SEIFA IN EXPLAINING DIFFERENCES IN HEALTH DETERMINANTS

In this section, we show by way of examples how SEIFA can be used to analyse the prevalence of health risk factors so that a targeted policy intervention can be implemented. Data from the National Health Survey (NHS) 2004–05 have been used in this analysis. More details on the survey can be found in *National Health Survey, 2004–05 – Summary of Results*, ABS cat. no. 4364.0, ABS (2006). The NHS 2004–05 collected information from about 26,000 individuals across Australia.

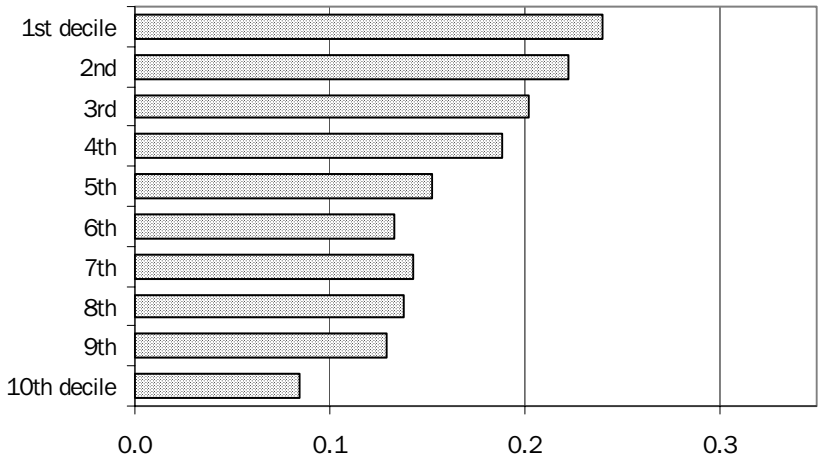
6.1 Self-reported health status

A general indicator of health is a person's self-assessment of their own health. There are examples of cultural differences on how a person perceives their own health, but generally this is considered a reliable measure of health status. Information on self-reported health status was collected in the NHS 2004–05 via the following question – 'In general would you say that your health is excellent, very good, good or fair?' The question was asked only for those respondents who were 15 years or over. Respondents who reported their health to be 'Fair' or 'Poor' were combined together to represent not-so-good health category. Just over 17% of respondents (3591 out of 20780 respondents aged 15 years and over) were in the 'Fair' or 'Poor' health category. The weighted prevalence of self reported health as being 'Fair' or 'Poor' among Australians aged 15 years and over was 16.1% (CI 15.5–16.7%).

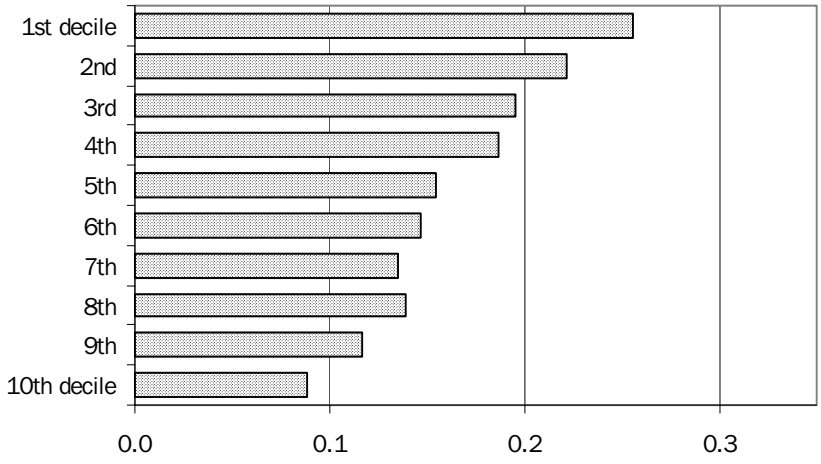
When only those respondents reporting 'Fair' or 'Poor' health status were examined further in terms of where they lived based on the SEIFA scores, the analysis showed that areas that were relatively more disadvantaged had a higher proportion of people with 'Fair' or 'Poor' health status (figure 6.1a). The proportion of respondents with 'Fair' or 'Poor' health status in areas with the lowest SEIFA decile (first decile) was three times as high as that of the area with the highest SEIFA decile (10th decile). We also notice that there is a gradient in the proportion with not so good health status – as the areas become relatively less disadvantaged, the proportion of people with not so good health status decreases. We also see that the gradient in health status – as the area becomes more disadvantaged, more people report their health as poor or fair – is evident for IRSAD as well (figure 6.1b).

6.1 Proportion of persons aged 15 years and over with Poor or Fair health, by SEIFA decile

(a) by SEIFA IRSD



(b) by SEIFA IRSAD



Source: *National Health Survey*, 2004–05

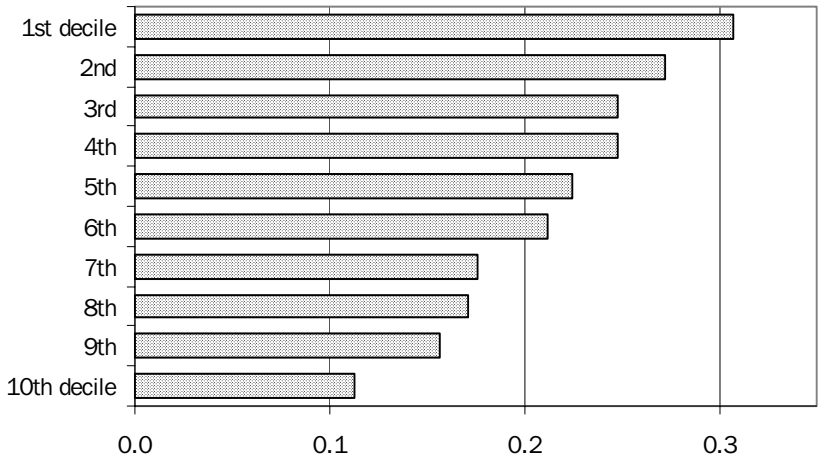
6.2 Smoking status

The NHS 2004–05 also asked questions about smoking for adult respondents (aged 18 years and over). There were a number of questions related to smoking ranging from age of initiation to current smoking habits. To identify whether anyone was a regular smoker, all respondents who answered 'Yes' to the question 'Do you currently smoke?' were asked a follow-up question 'Do you smoke regularly, that is, at least once a day?' Those reporting 'Yes' to this question were classified as a 'current daily smoker'. In all, 4231 respondents (out of 4585 all current smokers) reported that they currently smoked at least once a day. The weighted prevalence of daily smoking among Australian adults (aged 18 years and over) was 21.0% (CI 20.3–21.7%).

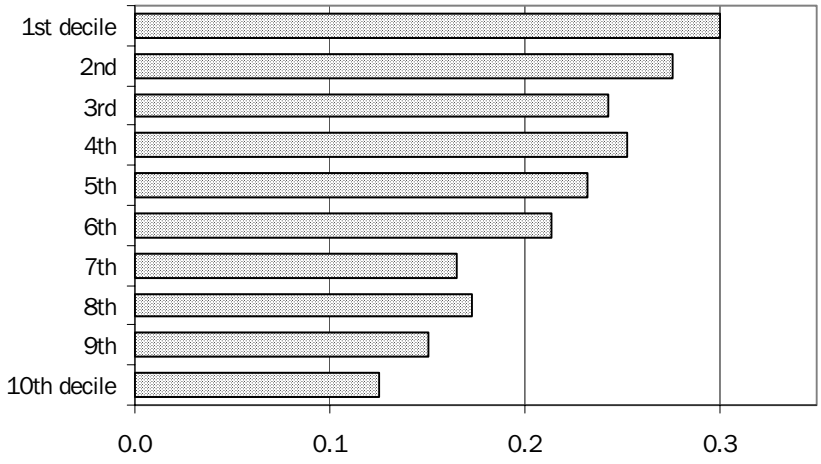
When prevalence of daily smoking was further analysed by SEIFA deciles, the results in figure 6.2a shows that, proportionately there were three times more daily smokers in areas which are relatively most disadvantaged (areas with first IRSD decile) compared to the least disadvantaged areas (i.e. areas in the 10th decile). The same pattern was observed when the prevalence of daily smoking was analysed by IRSAD (figure 6.2b).

6.2 Proportion of persons aged 18 years and over who are daily smokers, by SEIFA decile

(a) by SEIFA IRSD



(b) by SEIFA IRSAD



Source: *National Health Survey*, 2004–05

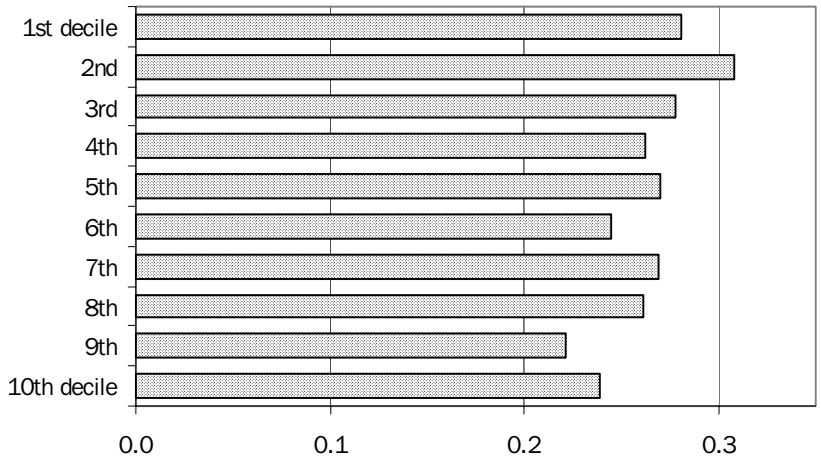
6.3 Prevalence of risky drinking

The NHS 2004–05 included a very detailed set of questions on alcohol consumption. As alcohol affects men and women differently, the risk level of alcohol for men and women is also different. A detailed question on graduated frequency level of alcohol consumption collected information on respondents' pattern and amount of alcohol consumption in a week. Using this information, alcohol risk levels – three-day average and seven-day average (using 2000 guidelines) have been established. In the following graphs (figures 6.3a and 6.3b) we have used the alcohol risk level based on the three-day average.

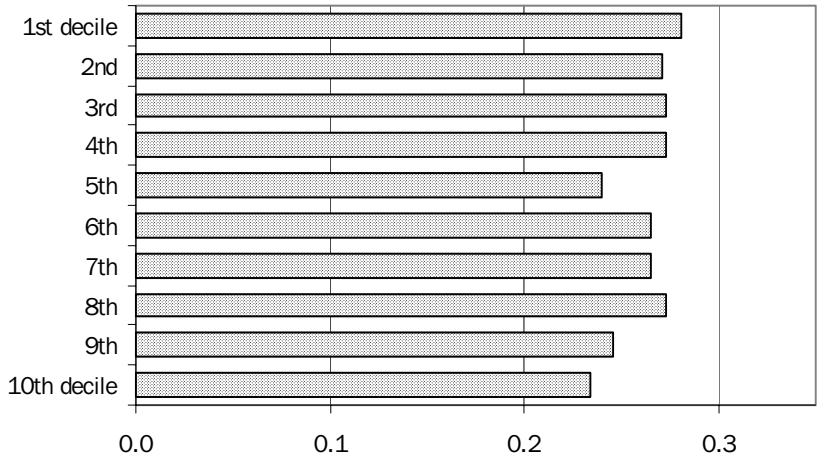
The three-day average alcohol risk level was divided into three risk categories – low risk, medium risk and high risk. Risk levels for the 12116 respondents could be established from the survey, with 5917 adult respondents falling in 'Low risk', 3073 in 'Medium risk' and 3216 in 'High risk' level. The weighted prevalence of high risk alcohol consumption among Australian adults (aged 18 years and over) was 26.1% (CI 25.1–27.0%). When we analysed high risk drinking behaviour by areas with SEIFA decile scores, the results showed that there was no observable gradient – there was no association with the proportion of high risk drinking behaviour by type of areas respondents lived (figures 6.3a and 6.3b).

6.3 Proportion of persons aged 18 years and over with high alcohol risk level, by SEIFA decile

(a) by SEIFA IRSD



(b) by SEIFA IRSAD



Source: *National Health Survey*, 2004–05

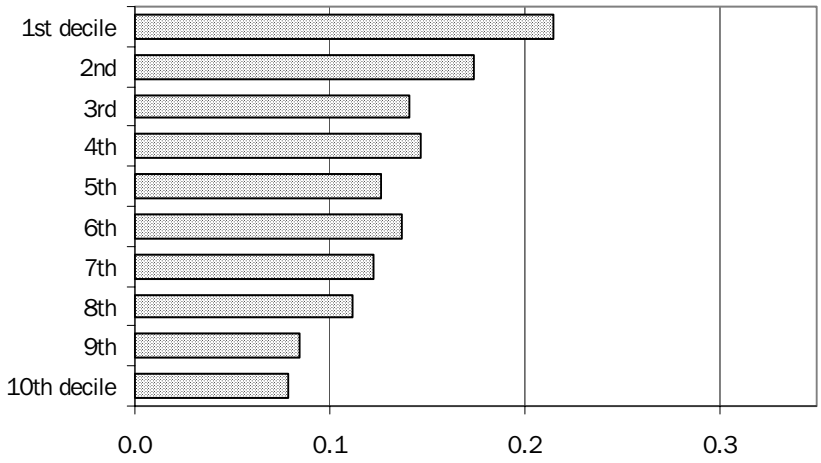
6.4 Self-reported mental health status

The NHS 2004–05 also included a set of 10 questions measuring respondents' self-reported mental health. The Kessler Psychological Distress scale (K10) was first documented by Kessler and Mroczek, of School of Survey Research Center, University of Michigan in 1994 (Kessler and Mroczek, 1994). The K10 Scale has been widely used in health surveys in Australia (Saunders and Daly, 2001; Dal Grande, Taylor and Wilson, 2000). Using the responses from 10 questions, grouped Kessler scores ranging from 1 to 4 were created for all respondents aged 18 years an over. In all, grouped Kessler scores for 19474 respondents were established, and of them 12261 were classified as 'Low distress level', 4603 as 'Moderate', 1826 as 'High' and 784 as 'Very high distress level'. In the analysis that follows, we have grouped respondents with 'High' and 'Very high' distress level into a single category. The weighted prevalence of 'High' or 'Very high' mental distress level among Australian adults (aged 18 years and over) was 13.1% (CI 12.6–13.8%).

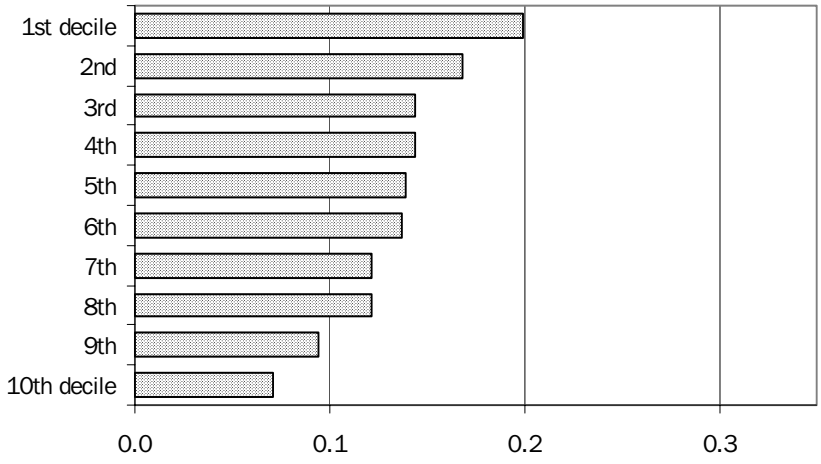
The analysis showed that, the lower an area scores on SEIFA IRSD index (relatively more disadvantaged), the higher is the proportion of Australians with high or very high distress level (figure 6.4a). The same pattern was observed for SEIFA IRSAD decile as well (figure 6.4b).

6.4 Proportion of persons aged 18 years and over with high distress level, by SEIFA decile

(a) by SEIFA IRSD



(b) by SEIFA IRSAD



Source: *National Health Survey*, 2004–05

6.5 Body Mass Index

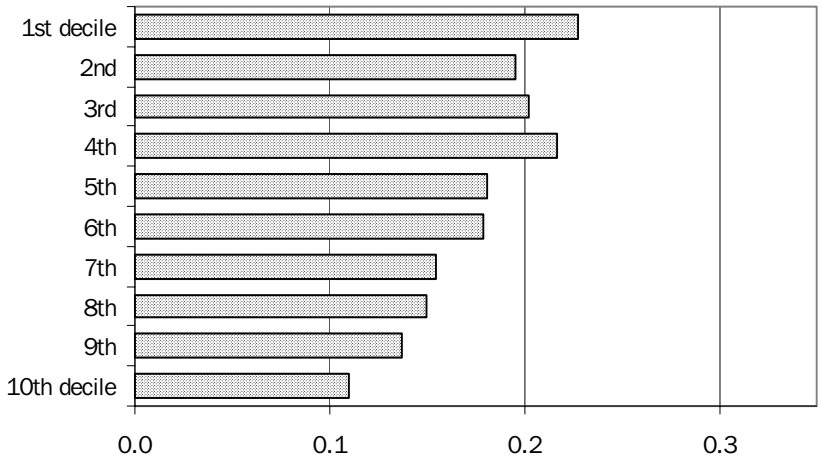
The NHS 2004–05 survey also asked respondents their weight and height. With this information, a Body Mass Index (BMI) was calculated for each respondent. A healthy BMI range is 18.5 to 25.0. A BMI in the range of 25–30 is considered overweight, and a BMI over 30 is considered obese. A low BMI is also not healthy and BMI less than 18.5 is considered underweight.

The BMI information was available for 18,744 respondents. Of them, 52% reported their BMI to be greater than 25. More than one in six respondents (3354 out of 18744) reported their BMI in the obese category (a BMI greater than 30). The overall mean BMI was 25.8 (CI 25.7–25.9), which is slightly higher than the healthy range. In the following analysis we only show the prevalence of obese people by SEIFA scores (figures 6.5a and 6.5b). The weighted prevalence of obesity among Australians aged 15 years and over was 17.3% (CI 16.7–18.0%).

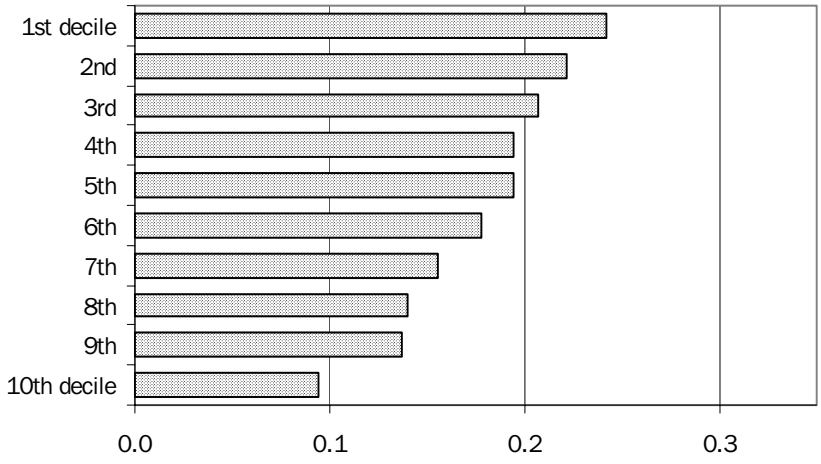
The figures 6.5a and 6.5b show that as the area's relative level of disadvantage increases, the proportion of people who are obese also increases. The point we have made earlier about the choice of index is evident here. If we look at the IRSD decile (figure 6.5a), we find that, although relatively more disadvantaged areas have proportionately more people who are obese, the pattern is not than linear compared to the gradient for the IRSAD decile (figure 6.5b).

6.5 Proportion of persons aged 15 years and over who are obese, by SEIFA decile

(a) by SEIFA IRSD



(b) by SEIFA IRSAD



Source: *National Health Survey*, 2004–05

In this section we demonstrated one of the uses of SEIFA scores. We could perform similar analysis on respondents' income, or any other attributes. We reiterate that a single SEIFA index may not be appropriate for all kind of analysis.

We saw the gradients in health determinants by SEIFA decile scores. Proportionately more people living in areas that are relatively more disadvantaged were found to be in poor health, more likely to be smoking, more likely to have higher mental stress and were also more likely to be obese. This is consistent with other research findings where area level socio-economic status has been found to be associated with population health (Nakaya and Dorling, 2005; Krieger et al., 2003). This shows that SEIFA scores which are summary measures of advantage/disadvantage indicator variables capture the construct that we have defined as 'disadvantage' and can explain the health differentials of people in areas that are relatively more disadvantaged.

7. AGGREGATION VERSUS DISTRIBUTIONAL ANALYSIS

So far we have shown the analysis based on the CD level SEIFA scores. Often, researchers want SEIFA scores at a more aggregated geographical areas either because their data do not contain CD level information, or because the number of cases in the analysis at the CD level is small. Aggregation of SEIFA scores to larger geographical areas often masks the relative difference in disadvantage that is present at the CD level. In the following examples we show the association between some of the health measures with the aggregated SEIFA scores and compare and contrast the gradient with the CD level score to show that SEIFA scores at aggregated geographic areas could mask the subtle difference that is present at the CD level.

We present three set of graphs: SEIFA deciles at CD level, CD level scores aggregated to SLA, and CD level scores aggregated to LGA level. To obtain average SEIFA scores at higher geographical levels, we first obtained the population weighted SEIFA scores (using CD population as weights), and then the average was estimated by dividing the sum of weighted scores by the total population of the SLA (to obtain SEIFA for SLA) or by LGA (to obtain SEIFA for LGA).

7.1 Self-reported health status

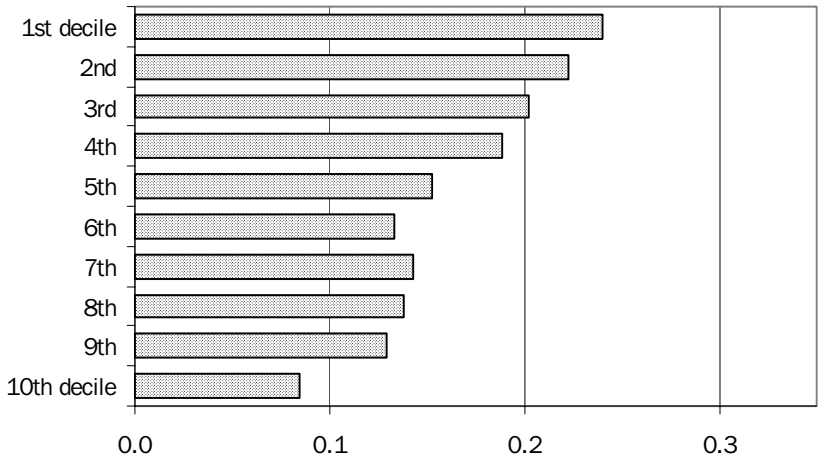
We have seen the first graph already (figure 7.1a). This is based on the decile scores at CD level. The second graph uses the IRSD deciles obtained by aggregating CD level SEIFA score to the SLA level (figure 7.1b). And in the third graph, we have created SEIFA deciles by aggregating the CD level scores to the LGA level (figure 7.1c). It is evident for self-reported health that the gradient that can be seen at the CD level analysis becomes less pronounced as we move from CDs to larger geographic areas (figures 7.1b and 7.1c).

7.2 Smoking status

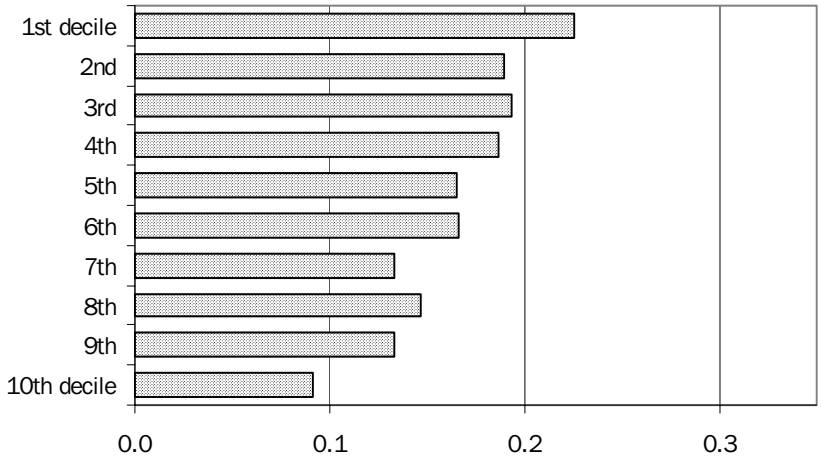
Similar analysis on prevalence of smoking by aggregated SEIFA IRSD deciles shows that the gradient that is present in the CD level decile (figure 7.2a) slowly disappears as we move from CD to SLA (figure 7.2b) and to LGA level deciles (figure 7.2c).

7.1 Proportion of persons aged 15 years and over with Poor or Fair health, by SEIFA

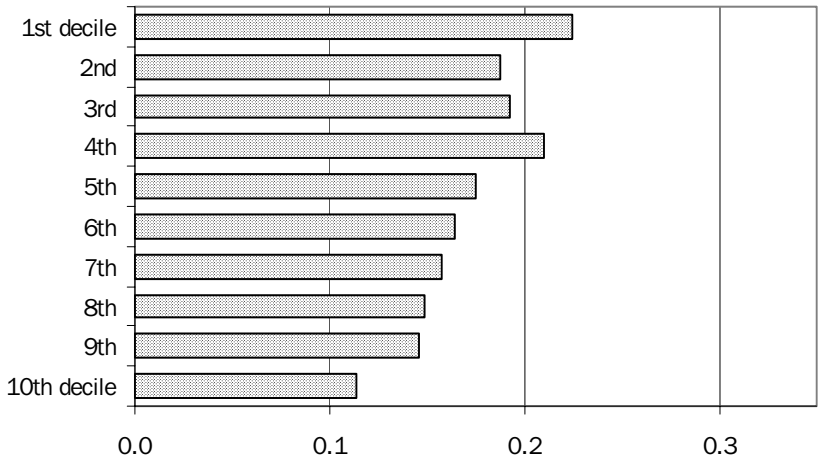
(a) CD level



(b) Aggregated to SLA



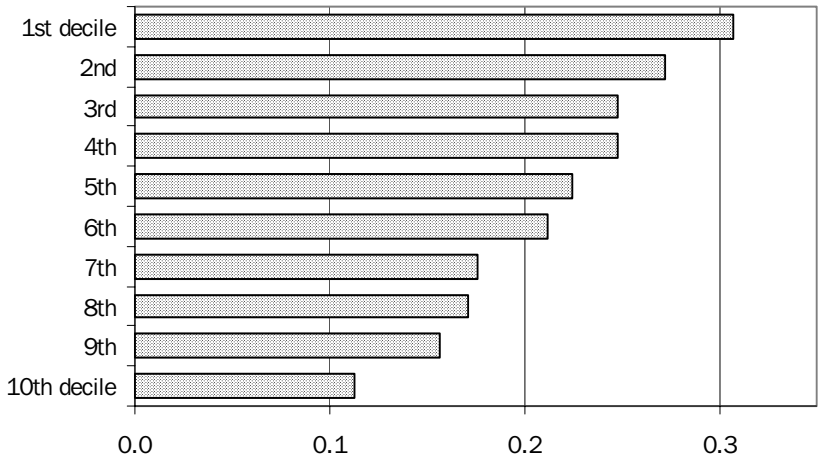
(c) Aggregated to LGA



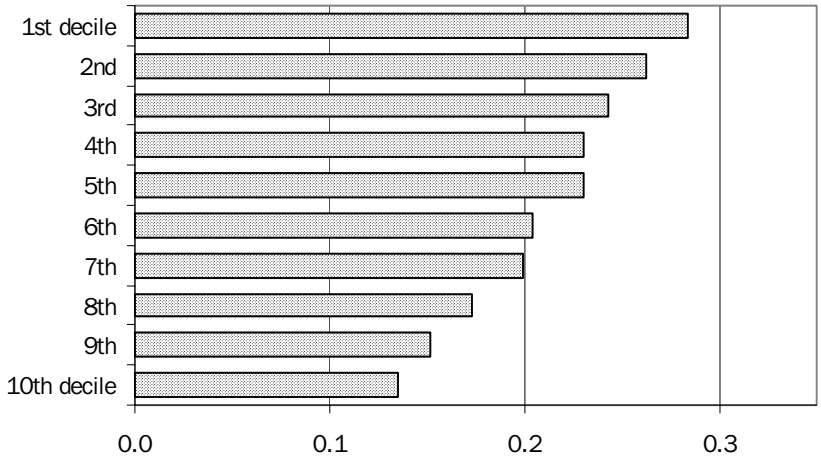
Source: *National Health Survey*, 2004–05

7.2 Proportion of persons aged 18 years and over who smoke regularly, by SEIFA

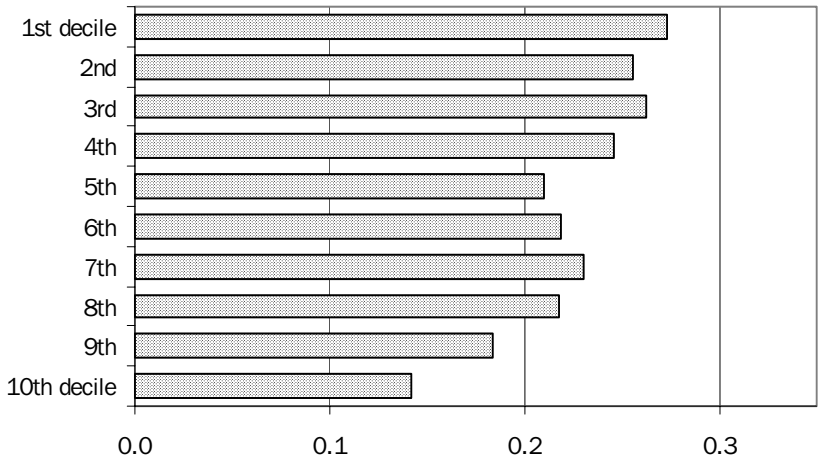
(a) CD level



(b) Aggregated to SLA



(c) Aggregated to LGA



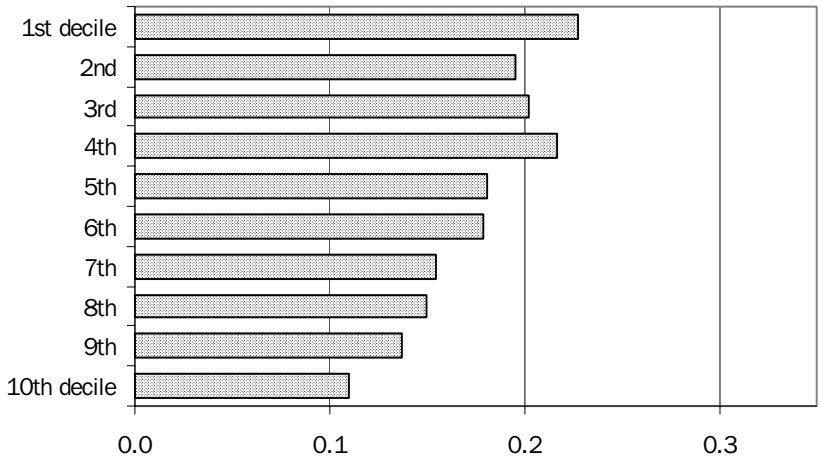
Source: *National Health Survey*, 2004–05

7.3 Prevalence of obesity

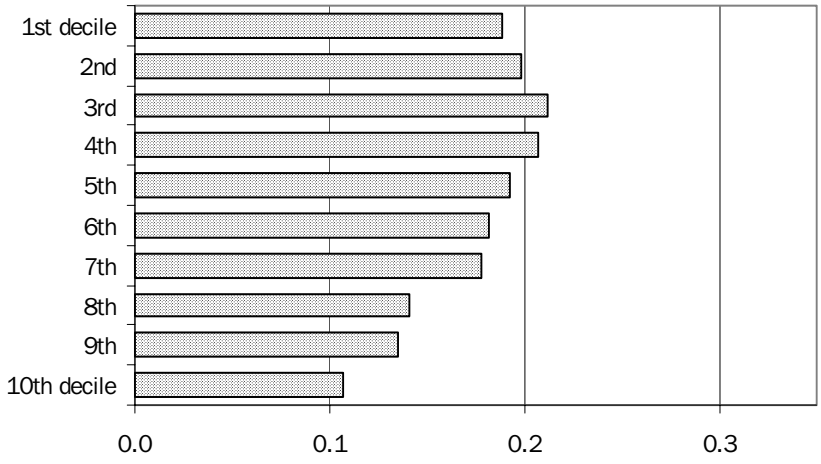
We saw earlier that there was a gradient in the prevalence of obese people by IRSD decile. The same graph is reproduced below (figure 7.3a). Although the gradient for the prevalence of obesity is not as linear as for self-reported health status or psychological distress level or the prevalence of daily smoking, we can still see that as the area become less disadvantaged (higher SEIFA decile), the prevalence rate of obesity decreases. As we move from CD level to SLA (figure 7.3b) and to LGA level analysis (figure 7.3c), the gradient in prevalence of obesity by IRSD decile takes a different shape. When we perform aggregated analysis, we see that for areas in either end of IRSD decile (1st or 10th deciles), the prevalence of obesity is lower compared to areas that are in 2nd, 3rd or 4th deciles (figure 7.3c).

7.3 Proportion of persons aged 15 years and over who are obese, by SEIFA

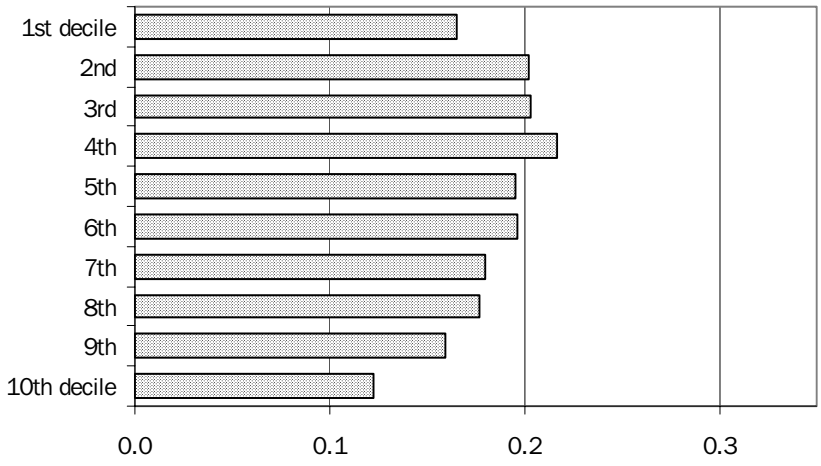
(a) CD level



(b) Aggregated to SLA



(c) Aggregated to LGA



Source: *National Health Survey*, 2004–05

In this section we highlighted the potential loss of information when we use aggregated SEIFA scores. As we have explained in earlier sections, aggregation has its own limitation. When we aggregate CD level scores to SLA, or to LGA or to any other larger geographic levels, and attempt to use this aggregated score to explain personal characteristics, any correlation we might observe is prone to ecological fallacy. We encourage researchers to use SEIFA at the lowest geographical level possible. In our examples we used a large national survey with a sample size of around 26,000. This large sample size allowed us to perform detailed analysis; using SEIFA deciles at CD level. There may be instances where researchers have to use aggregated analysis, or make the decile scores to quintile or quartile depending upon the sample size. We strongly encourage users to conduct distributional analysis to augment results obtained from aggregated analysis to avoid any ecological fallacy

8. PLANS FOR SEIFA 2006

For the SEIFA 2006, ABS is planning to release all the four indexes that were available in 2001. The same methodology will be used to summarise the Census variables. However, in 2006, a new disadvantage measure may be included. For the first time, a disability measure is included on the Census form. As we have explained in an earlier section, a person who cannot fully participate in social and economic activities is considered relatively more disadvantaged than their able bodied counterparts. We therefore plan to utilise this extra variable that is available to construct the SEIFA indexes. In terms of estimating SEIFA scores for other geographical areas (larger than CDs), we will separately estimate the principal component scores at these geography levels and provide appropriate SEIFA deciles. This is slightly different to what we did for SEIFA 2001. In 2001, we aggregated CD level SEIFA scores and calculated weighted average scores for geographical areas based on the CD scores.

The geographic levels for which separate scores (in decile form) will be available are: SLA, LGA, Postal area, and Commonwealth Electoral level. Although SEIFA scores will be available for these geographical areas, we strongly encourage users to look at the distribution of scores at CD level within each larger geographical area to see how the scores are distributed.

The 2006 SEIFA scores will be available free from the ABS website. A detailed publication containing the methodology and how SEIFA scores can be used as a research tool will be released at the same time as the SEIFA 2006. We are also exploring some possibilities of web-based analysis of SEIFA which allow users to undertake distributional analysis and compare relative disadvantage level across or within large geographical areas.

The SEIFA 2006 will be available by March 2008. As the ABS is moving towards meshblock as the smallest geographical unit at which aggregated data could be made available, we will be looking at a release of meshblock SEIFA after the initial release in 2008. However, as meshblocks contain about 30 dwellings, we are reviewing our method and the list of variables so that SEIFA at meshblock level is meaningful.

ACKNOWLEDGEMENTS

The author acknowledges Marion McEwin, Alanna Sutcliffe and Peter Rossiter for their constructive comments and suggestions.

REFERENCES

Barnett, J.R. (2000) 'Does place of residence matter? Contextual effects and smoking in Christchurch', *New Zealand Medical Journal*, 113, pp. 433–435.

Central Bureau of Statistics (2003) *Characterization and Classification of Geographical Units by the Socio-Economic Level of the Population*, Pub. No. 13, Central Bureau of Statistics, Jerusalem, Israel.

Dal Grande, E.; Taylor, A. and Wilson, D. (2002) *South Australian Health and Wellbeing Survey, December 2000*, Centre for Population Studies in Epidemiology, Department of Human Services, South Australia.

Department of the Environment, Transport and Regions (2000) *Indices of Deprivation 2000*, Regeneration Research Summary No. 31, DETR, London.

Diez–Roux, A.V. (1998) 'Bringing context back into epidemiology: Variables and fallacies in multilevel analysis', *American Journal of Public Health*, 88(2), pp. 216–222.

Diez–Roux, A.V.; Nieto, F.J.; Muntaner, C.; Tyroler, H.A.; Comstock, G.W.; Shahar, E.; Cooper, L.S.; Watson, R.L. and Szklo, M. (1997) 'Neighborhood Environments and Coronary Heart Disease', *American Journal of Epidemiology*, 146, pp. 48–63.

Glover, J.; Rosman, D. and Tennant, S. (2004) 'Unpacking analyses relying on area-based data: are the assumptions supportable?', *International Journal of Health Geographics,* 3(30). Published online.

Glover, J.D., Hetzel, D.M.S. and Tennant, S.K. (2004) 'The socioeconomic gradient and chronic illness and associated risk factors in Australia', *Australia and New Zealand Health Policy*, 1(8). Published online.

Glover, J.; Harris, K. and Tennant, S. (1999) *A Social Health Atlas of Australia*, Public Health Information Development Unit, University of Adelaide.

Health Information Partnership – Eastern Ontario Region (2005) *Census 2001 Profile of Eastern Ontario: Index of Relative Socio-economic Disadvantage*, Research Report, February 2005.

Kessler, R. and Mroczek, D. (1994) *Final Versions of our Non-Specific Psychological Distress Scale*, University of Michigan, Michigan.

Krieger N. (2006) *Geocoding and Monitoring US Socioeconomic Inequalities in Health: An introduction to using area-based socioeconomic measures*, The Public Health Disparities Geocoding Project Monograph, (accessed 19/06/2006) <http://www.hsph.harvard.edu/thegeocodingproject/webpage/monograph/absms.htm>

Krieger, N.; Chen J.T.; Waterman, P.D.; Rehkopf, D.H. and Subramanian, S.V. (2003) 'Race/Ethnicity, Gender, and Monitoring Socioeconomic Gradients in Health: A Comparison of Area-Based Socioeconomic Measures – The Public Health Disparities Geocoding Project', *American Journal of Public Health*, 93(10), pp. 1655–1671.

Nakaya, T. and Dorling, D. (2005) 'Geographical inequalities of mortality by income in two developed island countries: A cross-national comparison of Britain and Japan', *Social Science and Medicine*, 60, pp. 2865–2875.

Ross, N.A.; Wolfson, M.C.; Dunn, J.R.; Berthelot, J.M.; Kaplan, G.A. and Lynch, J.W. (2000) 'Relationship between income inequality and mortality in Canada and in the United States: Cross sectional assessment using Census data and vital statistics, *British Medical Journal*, 320, pp. 898–902.

Salmond, C. and Crampton P. (2001) 'NZDEP96 – What does it measure?', *Social Policy Journal of New Zealand*, 17, pp. 82–100.

Salmond, C. and Crampton P. (2002) *NZDEP2001 Index of Deprivation*, Department of Public Health, Wellington School of Medicine and Health Sciences, Wellington

Saunders, D. and Daly, A. (2001) *2000 Collaborative Health and Wellbeing Survey: Psychological Distress in the Western Australia Population*, Health Department of Western Australia, June 2001.

Sorlie, P.D.; Backlund, E. and Keller, J.B. (1995) 'US Mortality by Economic, Demographic, and Social Characteristics: The National Longitudinal Mortality Study', *American Journal of Public Health*, 85(7), pp. 949–956.

Townsend, P. (1987) 'Deprivation', *Journal of Social Policy*, 16, pp. 125–146.

Walker, R. and Hiller, J.E. (2005) 'The Index of Relative Socio-economic Disadvantage: General population views on indicators used to determine area-based disadvantage', *Australian and New Zealand Journal of Public Health*, 29(5), pp. 442–447.

Yu, X.Q.; Robertson, C. and Brett, I. (2000) 'Socioeconomic correlates of mortality differentials by Local Government Area in rural northern New South Wales, 1981–1995', *Australian and New Zealand Journal of Public Health*, 24(4), pp. 365–369.

APPENDIX

A.1 List of variables used for the Index of Relative Socio-Economic Disadvantage and their weights

Variable	*Weight*
% Persons aged 15 years and over with no qualifications	–0.3052
% Families with offspring having parental income less than $15,600	–0.2927
% Females in labour force unemployed	–0.2750
% Males in labour force unemployed	–0.2702
% Employed females classified as 'Labourers and Related Workers'	–0.2689
% Employed males as classified as 'Labourers and Related Workers'	–0.2685
% One-parent families with dependent offspring only	–0.2536
% Persons aged 15 years and over who left school at Year 10 or lower	–0.2505
% Employed males classified as 'Intermediate Production and Transport Workers'	–0.2370
% Families with income less than $15,600	–0.2296
% Households renting from Government Authority	–0.2196
% Persons aged 15 years and over separated or divorced	–0.1949
% Dwellings with no motor car at dwelling	–0.1912
% Employed females classified as 'Intermediate Production and Transport Workers'	–0.1853
% Persons aged 15 years and over who did not go to school	–0.1848
% Indigenous	–0.1796
% Lacking fluency in English	–0.1468
% Employed females classified as 'Elementary Clerical, Sales and Service Workers'	–0.1342
% Occupied private dwellings with two or more families	–0.1279
% Employed males classified as 'Tradespersons'	–0.1131

A.2 List of variables used for the Index of Relative Socio-Economic Advantage/Disadvantage and their weights

Variable	*Weight*
% Persons aged 15 years and over with no qualifications	–0.2544
% Persons aged 15 years and over who left school at Year 11 or lower	–0.2445
% Couple families with no children with annual income less than $20,800	–0.2003
% Couple families with dependent child(ren) only with annual income less than $36,400	–0.1977
% Employed females classified as 'Labourers and Related Workers'	–0.1918
% Employed males classified as 'Labourers and Related Workers'	–0.1903
% Employed males classified as 'Intermediate Production and Transport Workers'	–0.1861
% Single person household with income less than $15,600	–0.1803
% Males (in labour force) unemployed	–0.1611
% Females (in labour force) unemployed	–0.1579
% Couple families with dependents and non-dependents or with non-dependents only with annual income less than $52,000	–0.1464
% One-parent families with dependent offspring only	–0.1292
% Employed males classified as 'Tradespersons'	–0.1274
% Employed females classified as 'Intermediate Production and Transport Workers'	–0.1248
% Employed females classified as 'Elementary Clerical, Sales and Service Workers'	–0.0988
% Single parent families with dependents and non-dependents or with non-dependents with annual income less than $26,000	–0.0986
% Dwellings with four or more bedrooms	0.0845
% Employed females classified as 'Advanced Clerical and service Workers'	0.0958
% Single parent families with dependents and non-dependents or with non-dependents with annual income greater than $62,399	0.1271
% Employed males classified as 'Associate Professionals'	0.1354
% Persons aged 15 years and over at university or other tertiary institution	0.1482
% Single parent families with dependent child(ren) only with annual income less than $15,600	0.1674
% Couple families with dependents and non-dependents or with non-dependents only with annual income greater than $103,999	0.1758
% Persons using Internet at home	0.1882
% Single person households with income greater than $36,399	0.1974
% Employed females classified as 'Professionals'	0.2086
% Persons aged 15 years and over having an advanced diploma or diploma qualification	0.2111
% Employed males classified as 'Professionals'	0.2269
% Couple families with no children with annual income greater than $77,999	0.2325
% Couple families with dependent child(ren) only with annual income greater than $77,999	0.2381
% Persons aged 15 years and over with degree or higher	0.2440

A.3 List of variables used for the Index of Economic Resources and their weights

Variable	*Weight*
% Couple families with dependent child(ren) only with annual income greater than $77,999	0.3264
% Couple families with no children with annual income greater than $77,999	0.3235
% Single person households with income greater than $36,399	0.2988
% Households paying rent greater than $225 per week	0.2951
% Households paying mortgage greater than $1,360 per month	0.2859
% Couple families with dependents and non-dependents or with non-dependents only with annual income greater than $103,999	0.2646
% Single parent families with dependent child(ren) only with annual income greater than $36,399	0.2419
% Single parent families with dependents and non-dependents or with non-dependents with annual income greater than $62,399	0.2025
% Dwellings with four or more bedrooms	0.1341
% Single parent families with dependents and non-dependents or with non-dependents with annual income less than $26,000	–0.1554
% Households paying rent less than $88 per week	–0.1905
% Couple families with dependents and non-dependents or with non-dependents only with annual income less than $52,000	–0.2254
% Single person household with income less than $15,600	–0.2725
% Couple families with no children with annual income less than $20,800	–0.2774
% Couple families with dependent child(ren) only with annual income less than $36,400	–0.2840

A.4 List of variables used for the Index of Education and Occupation and their weights

Variable	*Weight*
% Persons aged 15 years and over with degree or higher	0.3282
% Employed males classified as 'Professionals'	0.3118
% Employed females classified as 'Professionals'	0.2943
% Persons aged 15 years and over having an advanced diploma or diploma qualification	0.2780
% Persons aged 15 years and over at University or other tertiary institution	0.2121
% Employed males classified as 'Associate Professionals'	0.1822
% Employed males classified as 'Advanced Clerical and Service Workers'	0.1173
% Employed females classified as 'Elementary Clerical, Sales and Service Workers'	–0.1376
% Males (in Labour Force) unemployed	–0.1677
% Employed females classified as 'Intermediate Production and Transport Workers'	–0.1759
% Females (in Labour Force) unemployed	–0.1778
% Employed males classified as 'Tradespersons'	–0.1906
% Employed males classified as 'Labourers and Related Workers'	–0.2371
% Employed females classified as 'Labourers and Related Workers'	–0.2490
% Employed males classified as 'Intermediate Production and Transport Workers'	–0.2612
% Persons aged 15 years and over with no qualifications	–0.3188
% Persons aged 15 years and over who left school at Year 11 or lower	–0.3244

FOR MORE INFORMATION . . .

INTERNET	**www.abs.gov.au** the ABS web site is the best place for data from our publications and information about the ABS.
LIBRARY	A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our web site, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE	1300 135 070
EMAIL	client.services@abs.gov.au
FAX	1300 135 211
POST	Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All ABS statistics can be downloaded free of charge from the ABS web site.

WEB ADDRESS	www.abs.gov.au

2000001544808
ISBN 0 642 48252 7

RRP $11.00