1352.0.55.086

Australian Bureau of Statistics

Research Paper

Socio-Economic Indexes for Individuals and Families

www.abs.gov.au

New Issue

Research Paper

Socio-Economic Indexes for Individuals and Families

Joanne Baker and Pramod Adhikari

Analytical Services Branch

Methodology Advisory Committee

8 June 2007, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THURS 23 AUG 2007

ABS Catalogue no. 1352.0.55.086
ISBN 978 0 64248 365 2

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Ms Joanne Baker, Analytical Services Branch on Canberra (02) 6252 6992 or email <analytical.services@abs.gov.au>.

CONTENTS

SOCIO-ECONOMIC INDEXES FOR INDIVIDUALS AND FAMILIES

Joanne Baker & Pramod Adhikari
Analytical Services Branch

ABSTRACT

The Australian Bureau of Statistics has released Socio-Economic Indexes for Areas (SEIFA) based on the Census of Population and Housing since 1986. The SEIFA indexes are widely used measures of relative socio-economic status at a small area level. The indexes rank and identify areas that are relatively more, or less, disadvantaged. They provide contextual information about the area in which a person lives. Yet, within any area there will be individuals and sub-populations with very different characteristics to the overall population of the area. When we make judgments about individuals, based on the characteristics of the area in which they live, there is potential for error in our conclusions. This potential for error is referred to as the ecological fallacy.

Using Census data for Western Australia, this paper explores the feasibility of creating individual and family level socio-economic indexes using the same conceptual and methodological basis as SEIFA. The analysis shows that a feasible index of disadvantage for individuals and families can be created.

Both the individual and family level indexes showed a wide range of low index scores, reflecting a wide range of indicators of disadvantage and a high incidence of multiple disadvantage. However, we found a large amount of heaping on a small number of high index scores. These people, and families, experienced few or no indicators of disadvantage.

Using these indexes, we investigate the extent of the ecological fallacy when SEIFA is used as a proxy for individual and family level socio-economic status. The analysis shows that there is a large amount of heterogeneity in the socio-economic status of individuals and families within small areas. These findings indicate that there is a high risk of the ecological fallacy when SEIFA is used as a proxy for the socio-economic status of smaller groups within an area and there is considerable potential for misclassification error.

1. INTRODUCTION

The Australian Census of Population and Housing is a rich source of information on income, education, occupation, housing tenure and other characteristics which are associated with socio-economic status. After the 1971 Census, the Australian Bureau of Statistics (ABS) used this information to create a measure of socio-economic disadvantage. The ABS has released Socio-Economic Indexes for Areas (SEIFA) for each Census since 1986.

SEIFA is created at the Census Collection District (CD) level by summarising a range of area level Census variables which relate to a concept of access to material and social resources. The SEIFA indexes aim to identify and rank small areas that are relatively more, or less, disadvantaged. The SEIFA indexes are widely used in a range of research at the small area level. For individual and household level analyses, SEIFA can provide contextual information about the area in which a person lives. To support this type of analysis, the ABS includes a SEIFA measure on many of its public use confidentialised unit record files.

Although SEIFA is an area level measure, a literature review found that SEIFA is often used as a proxy for the socio-economic status of individuals. In this type of analysis all people within an area are assumed to have the same level of advantage or disadvantage. However, we know that within any area there will be individuals and sub-populations which have very different characteristics to the overall population of that area. For example, Kennedy and Firman (2004) showed that there were large differences when SEIFA scores were calculated separately for Indigenous and non-Indigenous populations in 483 areas throughout Queensland.

If we use area level data to make inferences about the characteristics of individuals, or subgroups within that area, our conclusions could potentially be misleading, or even wrong. The potential for this type of error is called the ecological fallacy. The ecological fallacy is most likely to be an issue in areas where the characteristics of individuals or other small groups are very different to the average characteristics of people in the area. Because of this type of issue, there is interest in the creation of an individual or family level index of relative socio-economic disadvantage from researchers, policy makers and the ABS.

This research paper describes an initial foray into the development of individual and family level indexes of relative socio-economic disadvantage using 2001 Census data for Western Australia. There were two main aims for this explorative work. The first is to improve understanding of SEIFA and its uses. The second is to stimulate discussion on how to create a socio-economic index for individuals from Census data.

The paper describes how we derived an individual level index (SEIFI) and a family level index (SEIFF) using a similar conceptual and methodological basis as is used for

SEIFA. Using these indexes we investigate the extent of the ecological fallacy when SEIFA is used as a proxy for individual and family level socio-economic status.

In the next section we describe the concept of disadvantage used for SEIFA and explore the differences between area level and individual level disadvantage. This section also looks at how the framework used for SEIFA can be adapted to the creation of individual and family level socio-economic indexes. We have included a discussion of some of the practical issues we found in adapting the Census variables from area level variables to the individual and family levels. This is followed, in Section 3, by a description of the data and methodology used for SEIFA and how this method can be adjusted to the development of individual and family level indexes. In Section 4 we see how well the indexes allow us to identify and rank individuals, and families, as more, or less disadvantaged. In Section 5 we use the new individual and family level indexes to investigate the ecological fallacy by analysing the heterogeneity of individuals and families within the same area. In Section 6 we summarise our findings and outline possible directions for future research into the construction of individual level indexes.

2. SOCIO-ECONOMIC INDEXES

2.1 The notion of relative socio-economic disadvantage [1]

Relative socio-economic disadvantage is a complex and multi-dimensional concept. Using a concept of relative socio-economic disadvantage means that we need to look at whether the conditions experienced by individuals, families, or subgroups can be considered deprived relative to the wider community (Townsend, 1987). Townsend (1979) described some of the dimensions of relative disadvantage in his definition of relative deprivation. Under this definition, an individual may be deprived "if they lack the material standards of diet, clothing, housing, household facilities, working, environmental and locational conditions and facilities which are ordinarily available in their society, and do not participate in or have access to the forms of employment, occupation, education, recreation and family and social activities and relationships which are commonly experienced or accepted" (page 413).

As an example of the multi-dimensional nature of relative disadvantage, consider a community with relatively high levels of material wealth. We could conclude that this community is relatively advantaged. But if this community also has very high crime rates, high unemployment, or experiences relatively high levels of pollution, the community could be considered relatively disadvantaged.

The Census only collects information on a few dimensions of relative disadvantage. This is a difficulty that often arises in identifying a measure of relative disadvantage which could cover many economic, social, physical and spiritual dimensions. Around the world numerous socio-economic indexes have been developed. Most of these indexes include at least three main characteristics: employment, education and financial well-being.

Based on this international research and the type of information collected during the Census, we define socio-economic disadvantage in terms of an individuals' access to material and social resources, and their ability to participate in society.

Area versus individual level disadvantage

Area level and individual level socio-economic disadvantage are two separate, though interrelated, concepts. There are a wide range of factors and concepts associated with both area and individual level disadvantage. There are also many interlinkages between the two. For the purposes of this paper we have decided to use working definitions of area level disadvantage and individual level disadvantage. These are discussed below.

1 The first part of this section is based on Adhikari (2006), page 5.

Area level disadvantage is related to the characteristics of the community or neighbourhood as reflected in the attributes of the people living in that area. These characteristics may also be related to a lack of social and public resources, or characteristics which limit the access of residents to material resources or their ability to participate in society. More disadvantaged areas may lack employment opportunities, educational facilities, or transport infrastructure. There may also be an inadequate stock of housing, low levels of social capital, or high pollution and crime rates.

Individual level socio-economic disadvantage is a more personal concept relating to a person's own ability to access resources and participate in society. Individual disadvantage is related to a wide range of personal circumstances including personal and household income, educational background and qualification levels, employment status and occupation, health and disability, and family structure.

There will be interactions between area and individual level socio-economic disadvantage. For example, area level disadvantage can impact on the well-being of the residents of that area. There is a long history of research into the impact of area level disadvantage on individual outcomes including health and educational outcomes.[2] On the other hand, individual disadvantage will affect how well a person can take advantage of the services and opportunities available in the area where they live.

2.2 The ABS Index of Relative Socio-economic Disadvantage: Area level

2001 SEIFA is a set of four indexes designed to capture different aspects of relative socio-economic disadvantage at the small area level. The smallest area used to calculate SEIFA is the Census Collection District (CD) level. SEIFA is also available for other small areas, such as Statistical Local Areas (SLA) and Local Government Areas (LGA).

Literature reviews and user consultations indicate that the most commonly used SEIFA index is the Index of Relative Socio-Economic Disadvantage (IRSD). The IRSD was designed to be a general measure of relative socio-economic disadvantage at the area level. The variables included in the index are listed in table 2.1. The table also shows the index weights[3] which are applied to each variable.

Since this index only summarises variables that indicate disadvantage, a low IRSD score indicates that an area has a relatively large proportion of low income families, people with little training, or people working in relatively low skilled occupations.

2 On neighbourhood effects and health some examples are: Engles (1845); Kawachi (2006). On educational outcomes some examples are: Ginther, et al. (2000) and Borjas (1995); for effects in Australia, see Jensen and Seltzer (2000).

3 See Section 3 for more information on how these weights are derived.

2.1 Variables used for the Index of Relative Socio-Economic Disadvantage

Variable	*Weight*	*Prevalence (%)*
% People aged 15 years and over with no qualifications	–0.3052	56.6
% Families with offspring having parental income less than $15,600	–0.2927	7.5
% Females in labour force unemployed	–0.2750	6.6
% Males in labour force unemployed	–0.2702	8.0
% Employed females classified as 'Labourers and Related Workers'	–0.2689	7.2
% Employed males classified as 'Labourers and Related Workers'	–0.2685	10.2
% One-parent families with dependent offspring only	–0.2536	8.8
% People aged 15 years and over who left school at Year 10 or lower	–0.2505	45.1
% Employed males classified as 'Intermediate Production and Transport Workers'	–0.2370	13.0
% Families with income less than $15,600	–0.2296	3.9
% Households renting from Government Authority	–0.2196	4.9
% People aged 15 years and over who are separated or divorced	–0.1949	10.8
% Dwellings with no motor car	–0.1912	10.6
% Employed females classified as 'Intermediate Production and Transport Workers'	–0.1853	2.5
% People aged 15 years and over who did not go to school	–0.1848	1.1
% Indigenous	–0.1796	2.2
% Lacking fluency in English	–0.1468	2.8
% Employed females classified as 'Elementary Clerical, Sales and Service Workers'	–0.1342	14.2
% Occupied private dwellings with two or more families	–0.1279	1.0
% Employed males classified as 'Tradespersons'	–0.1131	20.4

The low score suggests that this area is disadvantaged relative to other areas. Correspondingly, an area with a high index score is relatively less disadvantaged than other areas. It is important to note that a high score reflects lack of disadvantage. It does not necessarily mean that the area is relatively advantaged.

2.3 Individual and family level indexes

In the past, the ABS has calculated socio-economic indexes at an area level. The underlying concepts and methodology that we use to calculate area level indexes could also be applied to individual people, or families. In this paper we have decided to explore the derivation of an individual level index (SEIFI) and a family level index (SEIFF) using the same conceptual basis as the area level IRSD. However, there are some practical issues which need to be considered when creating indexes at the individual and family level from Census variables.

One practical issue is that many of the variables used in the IRSD focus on employment, education and current income. For some individuals, access to material and social resources, and the ability to participate in society will not be captured well by these variables. For some groups, particularly older people, access to resources and the ability to participate will be partly determined by factors such as wealth,

accumulated assets and health. Information on these factors is not collected in detail in the Census. Because of this, the creation of an index for older people based on Census variables may be somewhat problematic.

For other people, like children, access to resources and the ability to participate will be highly dependent on the socio-economic status of their parents, guardians and other family members. Because of these issues, in this initial exploratory work we have decided not to include people under the age of 15 or over the age of 64 in the calculation of our individual level index.

For practical purposes we will also assume that resources are shared equitably within families. So, couples are assumed to have the same access to material and social resources, and the same ability to participate in society. Similarly, we assume that children within a family have the same socio-economic status as their parents or guardians.

Individual and family level variables

The creation of the individual and family level indexes started with the same list of Census variables as shown in table 2.1. Each of these area level variables has been transformed into an individual and family level variable. For individuals, each area level variable is transformed into a binary variable. For example, the continuous area variable "% Occupied private dwellings with two or more families" becomes a binary variable taking the value 1 if the individual lives in an occupied private dwelling with two or more families, and 0 otherwise.

For the family level index, each of the dwelling and family level Census variables such as "% Occupied private dwellings with two or more families" or "% Families with income less than $15,600" are also transformed into binary variables.

There are also a wide range of person level Census variables such as unemployment, lack of qualifications, low education, occupation and Indigenous status. For these variables, the transformation from an area level variable into a family level variable is not so clear cut. This is because more than one family member can display the characteristic. For example, within a family there may be more than one unemployed person. Holding other factors – such as family structure and income – constant, the level of relative disadvantage for this family is likely to be higher when there are more unemployed people in the family. For simplicity in this initial investigation, we have decided to use binary variables for all family level indicators of disadvantage. Future work may investigate the use of family level variables which reflect how increasing prevalence within the family affects the family's level of relative disadvantage while also taking family structure into account.

Finally we considered the gender specific nature of the area level occupation and unemployment variables. In creating an index for individuals, the presence of these gender variables leads us to question whether relative socio-economic disadvantage differs by gender. There is also the question of whether the relationship between other variables and disadvantage may also vary by gender. Our early investigations found that there was little difference between the loadings on gender specific occupation and employment variables. The inclusion of a gender variable was also considered initially, but the variable failed to meet our inclusion criteria.[4] Because of these findings, we decided not to include gender specific variables in the calculation of our individual or family level indexes.

This leaves us with an initial set of 17 binary indicators which can be used in the next stage of the development process for our individual and family level indexes. The individual and family level variables are shown in table 2.2 along with the analogous area level variables. Table 2.2 also shows the percentage of individuals and families with each of these characteristics.

2.4 Excluded observations

Excluded from SEIFA

For consistency with SEIFA, our analysis only included people and families found in Western Australian Census Collection Districts (CDs) which were included in the original SEIFA analysis in 2001. CDs were excluded for reasons including small population size and low levels of response to variables used in SEIFA.

Excluded from the individual level analysis

For the individual level analysis, people were excluded if they did not respond to all person level, family level and dwelling level indicators. Due to the issues described in Section 2.2, we also excluded people under the age of 15 and people aged 65 years and over.

Excluded from the family level analysis

Families were excluded from the analysis if they were:

- in non-family or non-classifiable households (includes people in single person or group households)
- non private dwellings
- family or dwelling level indicators were missing for the family
- person level indicators were missing for at least one member of the family

4 See Section 3 for more information on the inclusion criteria.

After making these exclusions, we calculated the individual index using 915,429 people and calculated the family index using 384,350 families.

2.2 List of variables considered for the individual and family level indexes with prevalence

Areas	*Individuals*		*Families*	
Variables	*Variables*	*(%)*	*Variables*	*(%)*
% People aged 15 years and over with no qualifications	No qualifications	55.4	At least one member aged 15+ years with no qualifications	74.3
% People aged 15 years and over who left school at Year 10 or lower	Left school at Year 10 or lower	40.2	At least one member aged 15 years and over left school at Year 10 or lower	61.0
% Employed males classified as 'Tradespersons'	Employed as 'Tradesperson'	9.0	At least one member employed as 'Tradesperson'	16.3
% People aged 15+ years: separated or divorced	Separated or divorced	8.4	At least one member aged 15+ years: separated or divorced	15.0
% Employed females classified as 'Elementary Clerical, Sales and Service Workers'	Employed as 'Elementary Clerical, Sales and Service Worker'	6.9	At least one member employed as 'Elementary Clerical, Sales and Service Worker'	12.4
% Employed (males / females) as classified as 'Labourers and Related Workers'	Employed as 'Labourers and Related Worker'	5.9	At least one member employed as 'Labourers and Related Worker'	10.3
% One-parent families with dependent offspring only	Part of one-parent family with dependent offspring only	5.7	One-parent family with dependent offspring only	9.3
% Employed (males / females) as 'Intermediate Production and Transport Workers'	Employed as 'Intermediate Production and Transport Worker'	5.6	At least one member employed as 'Intermediate Production and Transport Worker'	10.5
% Families with income < $15,600	Family income < $15,600	5.2	Family income < $15,600	7.5
% (males / females) unemployed	Unemployed	4.8	At least one member unemployed	8.4
% Households renting from Government Authority	Household rents from Government Authority	3.7	Household rents from Government Authority	4.0
% Families with offspring: parental income < $15,600	Family has offspring and parental income < $15,600	3.3	Family has offspring and parental income < $15,600	4.5
% Indigenous	Indigenous	2.4	At least one member Indigenous	2.8
% Dwellings with no car at dwelling	Lives in dwelling with no car at dwelling	2.3	Lives in dwelling with no car at dwelling	3.1
% Occupied private dwellings with two or more families	Lives in occupied private dwelling with two or more families	1.9	Family lives in occupied private dwelling with two or more families	2.1
% Do not speak English well	Does not speak English well	1.5	At least one member does not speak English well	2.9
% People aged 15 years and over who did not go to school	Did not go to school	0.5	At least one member aged 15+ years did not go to school	1.1

3. PRINCIPAL COMPONENT ANALYSIS

3.1 The method

The SEIFA indexes are calculated using a technique called Principal Components Analysis (PCA). PCA is used to reduce a large number of related, or correlated, variables into a smaller set of transformed variables, called 'components'. The components capture much of the information, or variation, contained in the original variables.

The first principal component accounts for the largest proportion of the variation in the original data set. The rest of the principal components are extracted so that they are uncorrelated with each other and account for progressively 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 summarise a large number of related variables into a small number of meaningful components. IRSD is the first principal component created from a set of 20 variables which indicate disadvantage in an area. 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).

Results from the PCA include:

- *Loadings:* which indicate the relationship, or correlation, between each of the observed variables and the principal components.
- *Eigenvalues:* which indicate how much variance in the original variables is explained by each component.
- *Weights:* which are calculated by dividing each loading by the square root of the eigenvalue.
- *Scores:* which are calculated by
    - standardising each of the original variables,
    - multiplying each standardised variable by the appropriate weight, and
    - summing to produce a raw score for each unit in the analysis (e.g. CD, person, family),
    - for presentation purposes, the raw score is standardised to a mean of 1,000 and standard deviation of 100.

PCA is usually based on a set of continuous variables – or a set of ordinal variables which are treated as if they are continuous. The correlation matrix for these variables is commonly calculated using Pearson's ρ. The SEIFA indexes were constructed using this type of PCA. If we use Pearson's ρ to calculate the correlation matrix for our

binary individual and family level variables, our PCA results will be biased (Rigdon and Ferguson, 1991).

Because of this, we have conducted PCA based on a tetrachoric correlation matrix. Tetrachoric correlation (or polychoric correlation for ordinal variables) calculates the correlation between latent variables which are assumed to underlie the binary variables. For example, although we only observe whether a person is unemployed or not unemployed, we assume that there is an underlying continuous variable which determines these two outcomes. The correlation matrices for individuals and for families are shown in Appendix A.

3.2 Creating the individual and family level indexes

Removal of highly correlated variables

Although PCA is based on the correlation of a set of variables, highly correlated variables may lead to instability in the PCA weights. So, before beginning our analysis, we needed to identify highly correlated variables and decide whether to drop one of the two variables. In line with the decision rule for IRSD, if the (tetrachoric) correlation coefficient of two variables was greater than 0.8 we consider the two variables to be highly correlated. For both individuals and families (see Appendix A for the correlation matrices), we found very high correlation between:

1. *Low family income* and *Low parental income*, and
2. *No schooling* and *Left school at year 10 or earlier*.

Low parental income is a subset of *Low family income*, and *No schooling* is a subset of having *Left school at year 10 or earlier*. So, we would expect to find a high correlation between these pairs of variables at the individual and family level. For these two pairs of variables, we decided to drop the two variables with lower prevalence: *Low parental income* and *No schooling*. The prevalence of these variables was shown in table 2.2.

For individuals, we also found very high negative correlation between being unemployed and each of the occupation variables. This is because unemployed people cannot be employed in any occupation and *vice versa*. The tetrachoric correlation matrix for individuals, shown in table A.1 of the Appendix, indicates that the occupation variables tend to have a negative correlation with the other indicators of disadvantage. This suggests that being employed, in any occupation, may not be a good indicator of disadvantage for individuals. Because of this, we decided to drop all the occupation variables from our analysis of individuals. This leaves us with 11 binary variables for the individual analysis and 15 variables for the family analysis. These variables are listed in table 3.1.

3.1 List of initial individual level variables

Individual level variables	*Family level variables*	*Code*
1. Does not speak English well	1. At least one member does not speak English well	*englishpoor*
2. Indigenous	2. At least one member Indigenous	*indigenous*
3. Family income < $15,600	3. Family income < $15,600	*lowincfam*
4. Lives in private dwelling with two or more families	4. Lives in private dwelling with two or more families	*multifam*
5. Lives in dwelling with no car at dwelling	5. Lives in dwelling with no car at dwelling	*nocar*
6. No qualifications	6. At least one member aged 15+ years with no qualifications	*noqual*
7. Part of one-parent family with dependent offspring only	7. Part of one-parent family with dependent offspring only	*oneparent*
8. Household rents from Government Authority	8. Household rents from Government Authority	*govrent*
9. Separated or divorced	9. At least one member aged 15+ years separated or divorced	*divorced*
10. Unemployed	10. At least one member unemployed	*unemp*
11. Left school at Year 10 or lower	11. Aged 15 years and over: left school at Year 10 or lower	*year10sch*
	12. At least one member employed as 'Elementary Clerical, Sales and Service Worker'	*cleric&sales*
	13. At least one member employed as 'Labourers and Related Worker'	*labourer*
	14. At least one member employed as 'Intermediate Production and Transport Worker'	*prod&trans*
	15. At least one member employed as 'Tradesperson'	*trades*

Removing variables poorly correlated with the first component

Now that we have identified the initial list of variables, we can undertake PCA using the tetrachoric correlation matrices. Because we are attempting to create indexes analogous to the 2001 index of disadvantage, we retained the first unrotated component for our family and individual indexes of disadvantage (see ABS, 2004, pp. 23–24 for more details). Although other components were considered, the first component seemed to provide the most intuitive index of disadvantage.

Once we have run our initial PCA we can look at the loading of each of the 11 variables for the first principal component. If a variable has a low loading, its weight in the index will normally be small. For IRSD, variables with a loading between –0.2 and 0.2 were dropped from the index. In the individual level analysis, we found that there were no variables with a loading of less than 0.2.

For the family level analysis, we found that having *Left school at year 10 or earlier* and being employed as a *Labourer* both had loadings between –0.2 and 0.2. We decided to drop both of these variables from our analysis. After dropping *Left school at year 10 or earlier* from the analysis, we decided to reintroduce *No schooling*. *No schooling* had previously been removed from the analysis due to high correlation between the two schooling variables.

At the family level, we also found that three of the occupation variables – *Tradesperson*, *Intermediate production and transport worker* and *Elementary clerical, sales and service worker* – had loadings of –0.47, –0.26 and –0.22 respectively. The relatively strong negative loading suggests that these variables are related to advantage rather than disadvantage. Since we only want to include variables which are related to disadvantage, we decided to drop variables with a negative loading.

Final loadings and weights

The final loadings for the family and individual PCA are shown in table 3.2. In both the individual and family level results, the variables with the highest loadings are *Living in a dwelling with no car* and being *Indigenous*. *Renting from a government authority* and *Living in a multifamily household* also had high loadings.

3.2 Loadings and weights for each of the indexes

	Individual		*Family*		*IRSD*	
Variable	*Loading*	*Weight*	*Loading*	*Weight*	*Loading*	*Weight*
nocar	0.80	0.44	0.75	0.42	0.49	0.19
indigenous	0.76	0.42	0.73	0.41	0.46	0.18
noschool	n/a	n/a	0.62	0.34	0.47	0.19
govrent	0.65	0.36	0.60	0.33	0.56	0.22
oneparent	0.56	0.31	0.47	0.26	0.65	0.25
multifam	0.54	0.30	0.56	0.31	0.33	0.13
lowincfam	0.53	0.30	0.53	0.29	0.59	0.23
noqual	0.44	0.25	0.40	0.22	0.78	0.31
year10sch	0.37	0.21	n/a	n/a	0.64	0.25
unemp	0.37	0.20	0.31	0.17	n/a	n/a
englishpoor	0.35	0.20	0.51	0.28	0.38	0.15
divorced	0.33	0.18	0.27	0.15	0.50	0.19

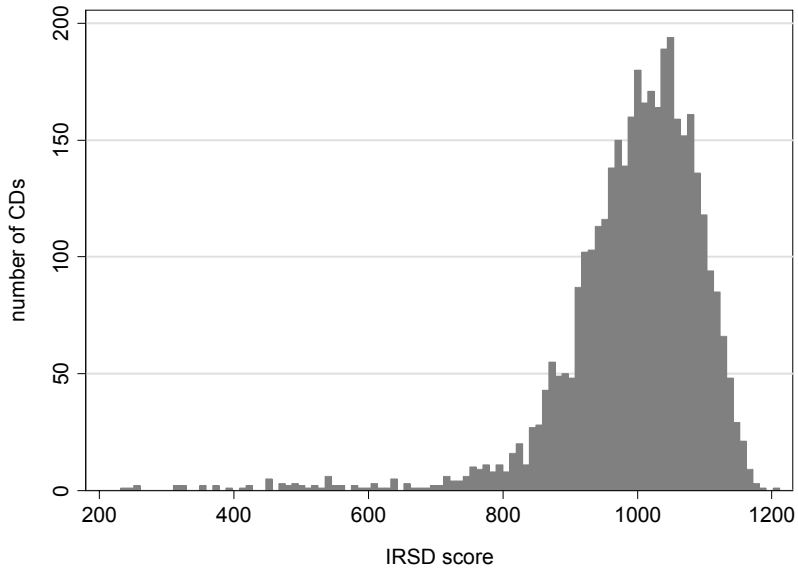
While most variables have similar loadings in both the family and individual analysis, *Not speaking English well* has a much higher loading for families (0.51) than for individuals (0.35). The two schooling variables – with only one variable included in each analysis – show large differences in loadings. *No schooling* has a loading of 0.62 in the family analysis, while *Left school at year 10 or earlier* has a loading of only 0.37 in the individual analysis.

3.3 The principal component scores

For each individual and for each family we can calculate a principal component score based on the weights given in table 3.2. In SEIFA, low scores indicate higher levels of disadvantage. For our individual and family level indexes we would also like low scores to represent higher levels of disadvantage. To achieve this, each of the weights in table 3.2 is multiplied by minus one. Then we follow the process outlined in Section 3.1 to calculate our individual and family scores. We standardise each of the original binary variables, multiply each standardised variable by the appropriate weight, and then sum to produce a raw score. For presentation purposes, the raw scores are standardised to a mean of 1,000 and standard deviation of 100.

As with SEIFA scores, both the socio-economic index for individuals (SEIFI) scores and the socio-economic index for families (SEIFF) scores are ordinal. For example, a family with a SEIFF score of 500 is not twice as disadvantaged as a family with a score of 1000.

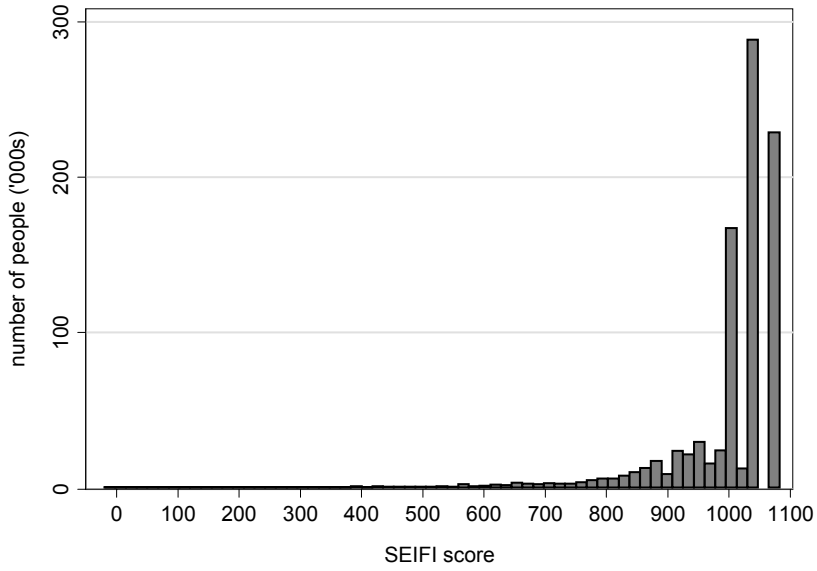
3.3 Distribution of IRSD scores



For comparison purposes, figure 3.3 shows the distribution of area level IRSD scores for CDs in Western Australia. Because IRSD only uses indicators of disadvantage there are few indicators to distinguish between CDs with relatively low levels of disadvantage. This results in scores which are skewed towards the bottom end of the distribution. The bottom 10% of scores range between 232 and 884. 80% of scores are within the range 884 to 1102 and the top 10% of scores only range between 1102 and 1211.

Figure 3.4 shows how the SEIFI scores are distributed across individuals. While the SEIFI scores range from a low of –20 to a high of 1075, the distribution is highly skewed towards the bottom end of the distribution, even when compared with the IRSD distribution. While there are a wide range of scores below 1002, there is a large amount of heaping on a few scores above 1002. One-quarter of scores are below 1002. The lowest 10% of scores were below 889, but only 1% of people have scores below 569.

3.4 Distribution of SEIFI scores



At the top end of the distribution we can see a large amount of heaping on particular scores. There are actually only five distinct scores above 1002. Table 3.5 shows the number of people with these top five scores and the indicators of disadvantage associated with each score. The top score, 1075, is given to all 228, 886 people who have no indicators of disadvantage. The next three highest scores are given to people with only 1 indicator of disadvantage. These people either have *No qualifications*, *Left school at year 10 or earlier*, or are *Separated or divorced*. The fifth highest score is given to people who have *Left school at year 10 or earlier* and also have *No qualifications*.

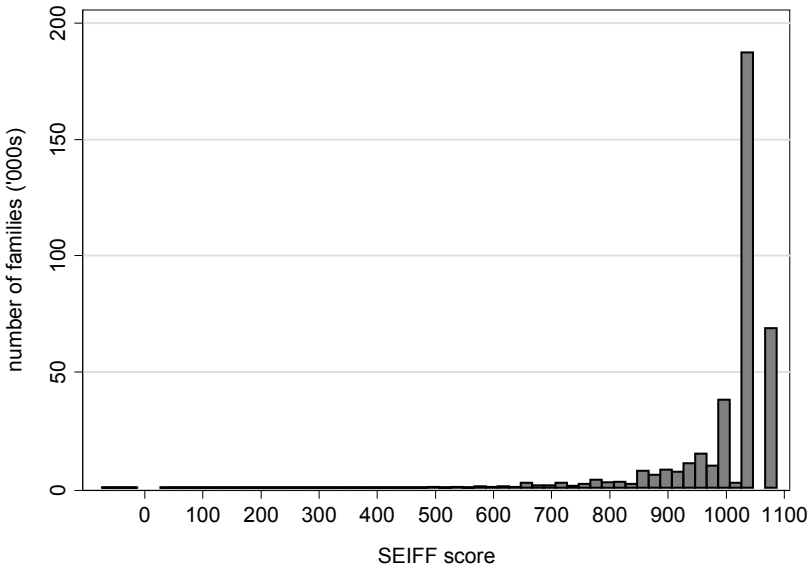
3.5 The top five SEIFI scores

SEIFI score	*Number of people*	*% of people*	*Indicators of disadvantage*
1004	160,133	17.5	*No qualifications* and *Left school at year 10*
1024	12,193	1.3	*Separated or divorced*
1037	193,853	21.2	*No qualifications*
1043	94,774	10.4	*Left school at year 10*
1075	228,886	25.0	*None*

No qualifications, *Left school at year 10 or earlier*, and being *Separated or divorced* are all indicators which have relatively low weightings (shown in table 3.2). They are also the three most prevalent of the eleven individual level indicators (see table 2.2). Since each binary variable is standardised to take account of prevalence, this results in higher scores for the most prevalent variables. The combination of high prevalence and relatively low weights result in high SEIFI scores for these people.

Figure 3.6 shows how the SEIFF scores are distributed across families. This distribution is very similar to the distribution of SEIFI scores. As with SEIFI scores, SEIFF scores are highly skewed towards the bottom end of the distribution. Again we see a wide range of low scores and a large amount of heaping on a few high scores. SEIFF scores range from a low of –73 to a high of 1077. Just over one-quarter of scores are below 1000. The lowest 10% of scores are below 887, but only 1% of families have scores below 569.

3.6 Distribution of SEIFF scores



There are only six distinct SEIFF scores above 1000. Table 3.7 shows the number of families with these top six scores and the indicators of disadvantage associated with each score. The top score, 1077, is given to all families with no indicators of disadvantage. Almost half of all families have at least one member with *No qualifications* and no other indicators of disadvantage. Each of these families is given a score of 1038. Other high scores are given to families with one relatively low weighting indicator of disadvantage. The sixth highest score is given to families who have at least one member who has *No qualifications* and one member who is *Separated or divorced*. As with the SEIFI scores, the combination of high prevalence and relatively low weights result in high SEIFF scores for these families.

3.7 The top six SEIFF scores

SEIFF score	*Number of families*	*% of families*	*Indicators of disadvantage*
1006	20,313	5.3	*No qualifications* and *Separated or divorced*
1008	2,254	0.6	*One parent family*
1030	3,244	0.8	*Unemployed*
1038	178,598	46.5	*No qualifications*
1045	5,619	1.5	*Separated or divorced*
1077	68,787	17.9	*None*

4. ANALYSING INDIVIDUAL AND FAMILY INDEXES

4.1 Creating SEIFI and SEIFF groups

To examine the extent of relative disadvantage, SEIFA scores are often ranked into deciles or quintiles. This provides us with a relatively simple way of comparing characteristics of areas at the extremes of the distribution. Ideally, we would also like to group the SEIFI and SEIFF scores in a similar way. By definition, each deciles should each contain 10% of people or families, and each quintile should contain 20%. However, the heaping at the top end of the SEIFI and SEIFF distributions make it difficult to create groups of an equal size. For example, how should we split the 47% of families with a score of 1038? Or the 25% of people with a score of 1075?

We have roughly divided the SEIFI scores into quartiles. In practice each of these four groups contain around 20–30% of people. We also attempted to split the SEIFF scores into four even groups, but the 47% of families with one score make this highly problematic. The scores were split into the following groups:

Group 1: families with the bottom 20% of scores

Group 2: families with scores between 960 and 1030 (14% of families)

Group 3: families with the 2nd and 3rd highest SEIFF scores (48% of families)

Group 4: families with no indicators of disadvantage (18% of families)

Table 4.1 shows selected details of each SEIFI and SEIFF group.

4.1 Distribution of SEIFI and SEIFF scores by group

	SEIFI scores for Individuals				*SEIFF scores for families*			
Group	*N*	*Percent*	*Min score*	*Max score*	*N*	*Percent*	*Min score*	*Max score*
1	225,590	24.6	–20	1001	78,021	20.3	–73	959
2	172,326	18.8	1004	1024	53,325	13.9	960	1030
3	288,627	31.6	1037	1043	184,217	47.9	1038	1045
4	228,886	25.0	1075	1075	68,787	17.9	1077	1077
Total	915,429	100.0	–20	1075	384,350	100.0	–73	1077

4.2 Characteristics of the SEIFI and SEIFF groups

In Section 3 we examined the indicators of disadvantage experienced by people and families with the highest SEIFI and SEIFF scores. We found that these people, and families, have either no indicators of disadvantage or only one, or two, low weight and high propensity indicators of disadvantage. In this section we use the groups described in Section 4.1 to explore the characteristics of people and families with lower scores.

Table 4.2 shows the number of indicators of disadvantage experienced by each person within the four SEIFI groups. The highest SEIFI group (group 4) contains all of the 228,886 people with no indicators of disadvantage. In contrast, 87% of people in the lowest SEIFI group (group 1) have at least two indicators of disadvantage and over 50% have at least three indicators of disadvantage. It should be noted that 97% of people with the lowest 10% of SEIFI scores have at least two indicators of disadvantage and 43% have at least four indicators of disadvantage.

4.2 Characteristics of the SEIFI groups

	Number of indicators of disadvantage							
SEIFI group	*0*	*1*	*2*	*3*	*4*	*5*	*6–9*	*N*
1	0.0	12.6	35.7	34.2	11.7	4.1	1.6	225,590
2	0.0	7.1	92.9	0.0	0.0	0.0	0.0	172,326
3	0.0	100.0	0.0	0.0	0.0	0.0	0.0	288,627
4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	228,886
Total	228,886	329,268	240,748	77,207	26,407	9,212	3,701	915,429

While many people in the lowest SEIFI group have multiple indicators of disadvantage, some people in this group have only one indicator. For example, we found that all people with *No car* at their dwelling are in the lowest SEIFI group. For 7% of these people having *No car* is their only indicator of disadvantage. Similarly, even if they only have one indicator of disadvantage, all people who are *Indigenous*, *Rent from a Government Authority*, are part of a *One parent family*, *Live in a dwelling with two or more families*, have *Low family income*, are *Unemployed*, or *Do not speak English well* are in the lowest SEIFI group. These people are assigned relatively low SEIFI scores, because their one indicator of disadvantage has a combination of low prevalence and a relatively high weight.

The lowest SEIFI group also contains 30% of people with *No qualification*, 30% of people who *Left school at year 10 or earlier* and 84% of people who are *Separated or divorced*. Each of these people experienced multiple indicators of disadvantage.

Table 4.3 shows the number of indicators of disadvantage, for each family by SEIFF group. By our definition, all families in the highest SEIFF group (group 4) have no indicators of disadvantage. However, 95% of families in the lowest SEIFF group (group 1) have at least two indicators of disadvantage and over a third of these families have at least four indicators of disadvantage.

4.3 Characteristics of the SEIFF groups

	Number of indicators of disadvantage							
SEIFF group	*0*	*1*	*2*	*3*	*4*	*5*	*6–9*	*N*
1	0.0	5.0	37.6	38.5	13.3	4.5	1.6	78,021
2	0.0	16.4	83.6	0.0	0.0	0.0	0.0	53,325
3	0.0	100.0	0.0	0.0	0.0	0.0	0.0	184,217
4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	68,787
Total	68,787	196,508	73,881	30,030	10,347	3,523	1,274	384,350

As with the lowest SEIFI group, many families in the lowest SEIFF group have multiple indicators of disadvantage. However, some of these families have only one indicator of disadvantage. Each of these indicators has a combination of low prevalence and a relatively high weight. All families with *No car* at the dwelling, who *Rent from a Government Authority*, live in a *Multi-family household*, at least one member is *Indigenous*, *Did not go to school*, or *Does not speak English well* are in the lowest SEIFF group, even if the family has only one indicator of disadvantage.

The lowest SEIFF group also contains two-thirds of *One parent families*, 88% of *Low income families*, 45% of families with members who are *Unemployed* or *Separated or divorced*, and 24% of families where at least one member has *No qualifications*. Each of these families experienced multiple indicators of disadvantage.

Figures 3.1 and 3.2 showed that there are a wide range of SEIFI and SEIFF scores at the bottom end of the SEIFI and SEIFF distributions. In this section we have found that the wide range of scores is due to the range of indicators of disadvantage experienced in the lowest SEIFI and SEIFF groups, and the high incidence of multiple indicators of disadvantage.

IRSD scores can be used to identify areas which are relatively more disadvantaged than other areas. Since the variables included in the SEIFI and SEIFF fit the notion of disadvantage given in Section 2.1, we should be able to use our SEIFI and SEIFF scores to identify which individuals, or families, are relatively more disadvantaged than

others. For people and families with low scores, the wide range of scores indicates that we will have fairly good discriminatory power in identifing and ranking individuals and familys as relatively more, or less, disadvantaged. However, the large amount of heaping on a few high scores means that we will be very limited in our ability to identify, or rank, individuals with relatively low levels of disadvantage.

5. THE ECOLOGICAL FALLACY

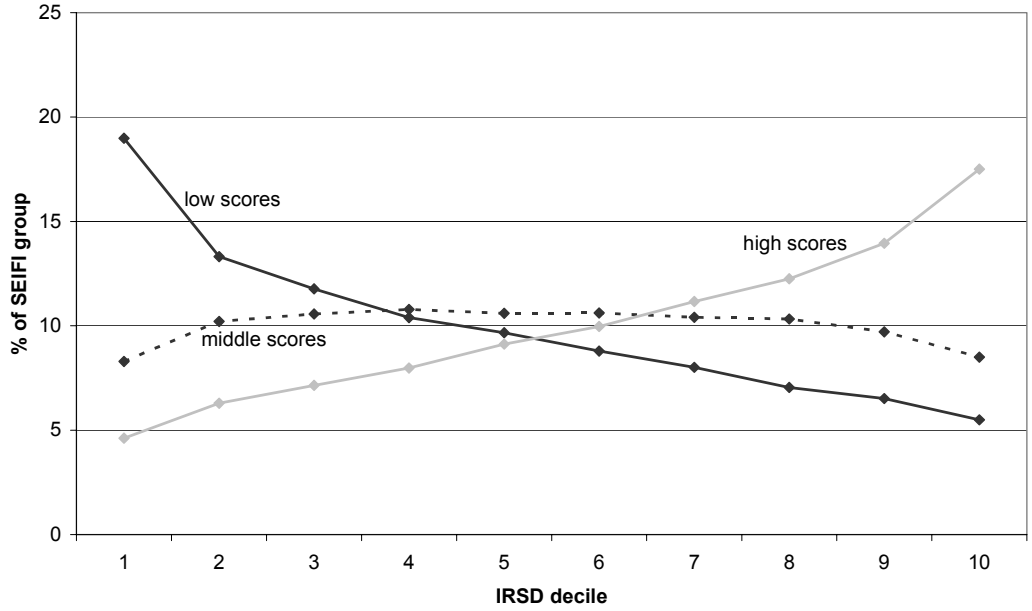
When there is no information available on the socio-economic status of individuals, an area level measure such as the SEIFA indexes is sometimes used as a proxy. This type of analysis assumes that all people in an area have the same socio-economic status. This assumption will not be valid if people within an area are heterogeneous in their characteristics and in their level of relative socio-economic disadvantage. There may be people living in a relatively more disadvantaged area who are not disadvantaged. In contrast, there may be people living in a relatively less disadvantaged area who are highly disadvantaged. If we use area level data, like the SEIFA scores, to make inferences about the characteristics of individuals, or subgroups within that area, our conclusions could potentially be misleading, or even wrong. The potential for this type of error is called the ecological fallacy.

The creation of SEIFF and SEIFI allows us to explore the extent of the ecological fallacy when the IRSD is used as a proxy for individual or family disadvantage. This can be determined by analysing the distribution of SEIFF and SEIFI scores within each of the IRSD deciles.

If there is a high level of homogeneity among people or households within each area, we will find a strong relationship between IRSD scores and both SEIFI and SEIFF scores. In the lowest IRSD decile we would expect to find a high level of disadvantage amongst the people and families residing in the area. Higher deciles are expected to have people and families who are relatively less disadvantaged than lower deciles. In this case there may be less risk of an ecological fallacy.

Figure 5.1 provides an illustration of how individuals in the SEIFI groups are distributed across the IRSD deciles. If SEIFI groups are distributed evenly across the IRSD decile, then we would expect to see around 10% of the SEIFI group in each IRSD decile. To simplify the graphic we have combined SEIFI groups 2 and 3. Appendix B contains more detailed information on the distribution of SEIFI and SEIFF scores across IRSD deciles.

5.1 Percent of people in each IRSD decile by SEIFI group

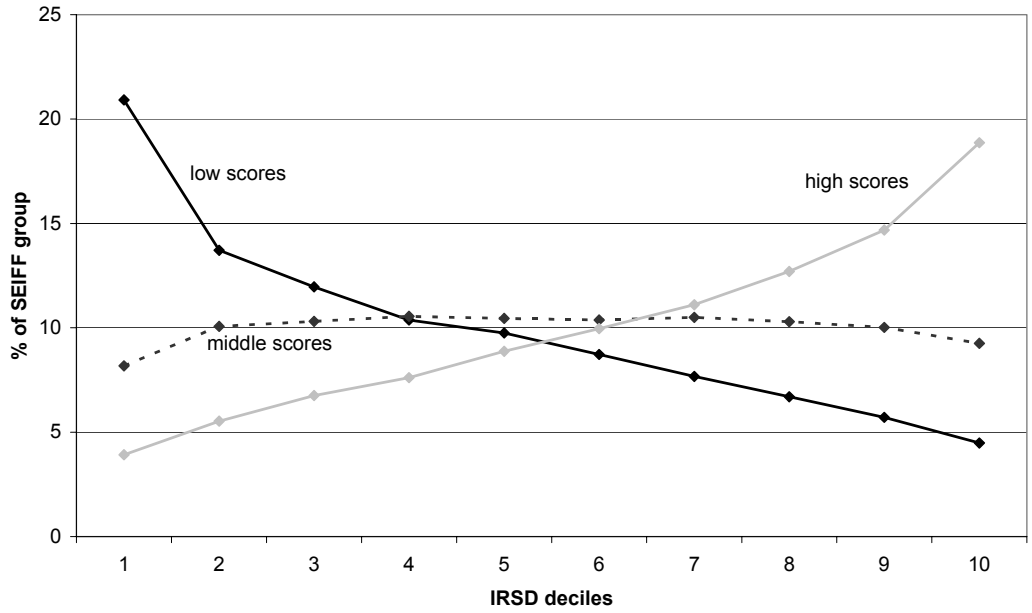


For the highest SEIFI group (who have no indicators of disadvantage) we can see a positive relationship with the IRSD deciles. Less than 5% of people in the highest SEIFI group live in the CDs of the lowest IRSD decile. This proportion rises with each IRSD decile, reaching 18% in the top IRSD decile. The reverse is seen for people in the lowest SEIFI group. 19% of people in the lowest SEIFI group live in CDs found in the lowest IRSD decile and less than 6% live in the CDs of the highest IRSD decile.

While there does appear to be a relationship between SEIFI and IRSD scores, over a third of people in the bottom SEIFI group live in the top five IRSD deciles. A similar proportion of people in the highest SEIFI group live in the bottom five IRSD deciles. We can also see in figure 5.1 that SEIFI groups 2 and 3 are fairly evenly distributed across the IRSD deciles.

Figure 5.2 shows similar patterns in the distribution of families across the IRSD deciles by SEIFF group. Again we can see a positive relationship between the IRSD deciles and the highest SEIFF group. We can also see a negative relationship with the lowest SEIFF group. However, as with SEIFI, around a third of families in the bottom SEIFF group live in the top five IRSD deciles and a similar proportion of the highest SEIFF group live in the bottom five IRSD deciles. SEIFF groups 2 and 3 are also fairly evenly distributed across each of the IRSD deciles.

5.2 Percent of families in each IRSD decile by SEIFF group



This analysis shows that using an area level indicator of socio-economic disadvantage will not be a good proxy for the socio-economic status of many of the individuals and families living within that area. Because of this, analyses which use SEIFA indexes such as the IRSD as a proxy for family and individual socio-economic status will be at high risk of an ecological fallacy.

6. CONCLUDING REMARKS

ABS has a long history of creating socio-economic indexes at an area level. In this research paper we presented the results of a preliminary exploration into the creation of individual and family level indexes of relative socio-economic disadvantage.

We found that the distribution of SEIFI and SEIFF scores were highly skewed towards the left. There were a wide range of low scores, reflecting a wide range of indicators of disadvantage and a high incidence of multiple disadvantage. At the top end of the distribution we found a large amount of heaping. These people, and families, experienced few or no indicators of disadvantage. The addition of indicators of advantage into the indexes may allow us to identify more and less advantaged individuals and families at the higher end of the distribution.

We used the individual and family indexes to examine whether there is a high risk of an ecological fallacy if the IRSD is used as a proxy for individual or family level disadvantage. Our analysis found that individual and family relative socio-economic disadvantage was quite diverse within areas. This means that there is a high risk of an ecological fallacy if we use the SEIFA indexes as a measure of individual level disadvantage, rather than a measure of area level disadvantage.

Comments from the ABS Methodology Advisory Committee

A version of this paper was presented to the ABS Methodology Advisory Committee (MAC) in June 2007. The MAC members were very enthusiastic about ABS working to create a socio-economic index for individuals. They encouraged ABS to continue with this development work, as they felt that this type of index would be very valuable for researchers and policy makers. MAC members maintained that area level indexes (i.e. SEIFA) are only used, incorrectly, as a proxy for individual socio-economic status because no other information is available. In addition to a census based individual index, MAC suggested that ABS should also consider an index that derived from variables included in social surveys. MAC acknowledged that future work on the development of this type of index needs to proceed carefully.

ABS is considering work to develop these indexes. Taking into account the findings from this preliminary work, and the comments from MAC, this would involve thorough investigation and resolution of a range of issues, including:

- A review of the definition of individual level disadvantage,
- The selection of the best individual level Census variables,
- The use of both advantage and disadvantage related variables,
- The minimisation of population exclusions,
- Indexes for different age groups,
- Validation process for the indexes.

User consultation would be an important part of any future development of individual and family level indexes of socio-economic disadvantage.

ACKNOWLEDGEMENTS

The authors would like to thank Marion McEwin, Jonathon Khoo, Clare Saunders, Nicholas Biddle, Jenny Myers, Peter Rossiter, the members of the SEIFA project board and participants of the June 2007 Methodology Advisory Committee Meeting for their helpful comments and assistance with this research project. The content and presentation of the paper are much improved as a result of their input. Responsibility for any errors or omissions remains solely with the authors.

BIBLIOGRAPHY

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

Adhikari, P. (2006) "Socio-Economic Indexes for Areas: Introduction, Use and Future Directions", *Methodology Research Papers*, cat. no. 1351.0.55.015, ABS, Canberra.

Borjas, G.J. (1995) "Ethnicity, Neighborhoods, and Human-Capital Externalities", *American Economic Review*, 85(3), pp. 365–390.

Darlington, R.B. (1997) *Factor Analysis* <http://www.psych.cornell.edu/Darlington/factor.htm> (last viewed 17/6/2007)

Engles, F. (1845 [1887]) *The Condition of the Working Class in England*, English translation <http://www.marxists.org/archive/marx/works/1845/condition-working-class/index.htm> (last viewed 07/07/2007)

Ginther, D.; Haveman, R. and Wolfe, B. (2000) "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust Are the Relationships?", *Journal of Human Resources,* 35(4), pp. 603–642.

Jensen, B. and Seltzer, A. (2000) "Neighbourhood and Family Effects in Educational Progress", *The Australian Economic Review*, 33(1), pp. 17–31.

Kennedy, B. and Firman, D. (2004) *Indigenous SEIFA – Revealing the Ecological Fallacy*, Paper presented to the 12th Biennial Conference of the Australian Population Association, 15–17 September 2004, Canberra.

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, <http://www.hsph.harvard.edu/thegeocodingproject/webpage/monograph/introduction.htm> (last viewed 07/07/2007)

Olsson, U. (1979) "Maximum Likelihood Estimation of the Polychoric Correlation Coefficient", *Psychometrika*, 44(4), pp.443–460.

Townsend, P. (1979) *Poverty in the United Kingdom, A Survey of Household Resources and Standards of Living*, London.

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

Rigdon, E.E. and Ferguson, C.E., Jr. (1991) "The Performance of the Polychoric Correlation Coefficient and Selected Fitting Functions in Confirmatory Factor Analysis with Ordinal Data", *Journal of Marketing Research*, 28, pp.491–497.

APPENDIXES

A. CORRELATION MATRICES

A.1 Tetrachoric correlation matrix for individual level index

	c&sales	*engpoor*	*indig*	*labourer*	*lowincf*	*lowincp*	*multifam*	*nocar*	*noqual*	*nosch*	*1parent*	*p&trans*	*govrent*	*divorced*	*trades*	*unemp*	*year10*
c&sales	1.00																
engpoor	–0.22	1.00															
indig	–0.12	0.08	1.00														
labourer	***–1.00***	0.19	0.22	1.00													
lowincf	–0.16	0.23	0.13	–0.11	1.00												
lowincp	–0.11	0.19	0.20	–0.07	***1.00***	1.00											
multifam	–0.04	0.31	0.51	0.14	0.23	0.23	1.00										
nocar	–0.12	0.25	0.64	0.13	0.34	0.36	0.34	1.00									
noqual	0.30	0.23	0.29	0.28	0.16	0.15	0.12	0.23	1.00								
nosch	–0.17	0.70	0.32	0.15	0.22	0.20	0.32	0.31	0.38	1.00							
1parent	0.04	–0.06	0.17	–0.04	0.36	0.53	0.22	0.37	0.14	–0.03	1.00						
p&trans	***–1.00***	–0.05	–0.07	***–0.92***	–0.29	–0.25	–0.01	–0.14	0.19	–0.06	–0.19	1.00					
govrent	–0.07	0.11	0.58	0.05	0.23	0.29	0.10	0.47	0.17	0.17	0.31	–0.08	1.00				
divorced	–0.04	–0.07	–0.02	0.00	0.12	0.23	0.02	0.16	0.00	–0.01	0.66	–0.02	0.17	1.00			
trades	***–1.00***	–0.07	–0.16	***–1.00***	–0.33	–0.29	–0.06	–0.23	–0.31	–0.09	–0.26	***–1.00***	–0.15	–0.08	1.00		
unemp	***–0.92***	0.09	0.17	***–0.93***	0.25	0.27	0.07	0.24	0.11	0.05	0.14	***–0.93***	0.18	0.08	***–1.00***	1.00	
year10	–0.07	0.19	0.26	0.17	0.12	0.04	0.12	0.18	0.29	***1.00***	–0.02	0.20	0.14	0.13	0.20	0.05	1.00

A.2 Tetrachoric correlation matrix for family level index

	c&sales	*engpoor*	*indig*	*labourer*	*lowincf*	*lowincp*	*multifam*	*nocar*	*noqual*	*nosch*	*1parent*	*p&trans*	*govrent*	*divorced*	*trades*	*unemp*	*year10*
c&sales	1.00																
engpoor	–0.11	1.00															
indig	–0.09	0.10	1.00														
labourer	0.01	0.17	0.20	1.00													
lowincf	–0.22	0.19	0.19	–0.17	1.00												
lowincp	–0.17	0.14	0.26	–0.13	***1.00***	1.00											
multifam	–0.07	0.33	0.46	0.10	0.29	0.28	1.00										
nocar	–0.19	0.31	0.55	0.03	0.33	0.33	0.25	1.00									
noqual	0.27	0.27	0.23	0.33	0.05	0.00	0.10	0.20	1.00								
nosch	–0.10	0.75	0.34	0.15	0.15	0.08	0.33	0.35	0.49	1.00							
1parent	–0.08	–0.14	0.28	–0.15	0.40	0.58	0.27	0.38	–0.11	–0.14	1.00						
p&trans	0.04	–0.03	–0.02	0.03	–0.36	–0.33	–0.04	–0.22	0.31	–0.04	–0.31	1.00					
govrent	–0.10	0.08	0.54	0.04	0.28	0.34	0.07	0.45	0.13	0.15	0.35	–0.11	1.00				
divorced	0.04	–0.11	0.05	0.02	0.07	0.21	0.01	0.13	0.01	–0.03	0.69	–0.04	0.19	1.00			
trades	0.03	–0.08	–0.13	–0.04	–0.42	–0.38	–0.10	–0.33	0.02	–0.12	–0.40	–0.17	–0.20	–0.10	1.00		
unemp	–0.06	0.10	0.16	–0.04	0.19	0.20	0.05	0.17	0.14	0.08	0.05	–0.09	0.15	0.18	–0.14	1.00	
year10	0.09	0.19	0.24	0.23	0.00	–0.07	0.05	0.14	0.51	***0.96***	–0.17	0.24	0.11	0.00	0.21	0.09	1.00

B. SEIFI AND SEIFF SCORES ACROSS IRSD DECILES

This appendix provides greater detail on the distribution of SEIFI and SEIFF scores across the IRSD deciles. Ideally, we would like to group the SEIFI and SEIFF scores into deciles, each containing 10% of scores. However, the heaping at the top end of the SEIFI and SEIFF distributions make it difficult to create groups of an equal size. In Section 5, figures 5.1 and 5.2 used three broad SEIFI and SEIFF groups. In this appendix we split the three broad groups into smaller groups which are described below.

The broad group labeled 'low' in Section 5 becomes:

- three smaller groups for SEIFI labeled "1" to "3", where smaller group "1" contains the lowest 10% of SEIFI scores, and
- two smaller groups for SEIFF labeled "1" and "2", where group "1" contains the lowest 10% of SEIFF scores

The broad group 'middle' becomes four smaller groups labeled "4" to "7".

The broad group 'high' remains as one group (containing all people and families with no indicators of disadvantage). This is labeled as smaller group "8".

Table B.1 provides details on the smaller SEIFI and SEIFF groups.

B.1 Distribution of SEIFI and SEIFF scores by group

		SEIFI scores for Individuals				*SEIFF scores for families*			
Broad group	*Smaller group*	*N*	*Percent*	*Min*	*Max*	*N*	*Percent*	*Min*	*Max*
Low	1	90,481	9.9	–20	883	38,350	10.0	–73	884
	2	88,989	9.7	889	954	39,671	10.3	887	959
	3	46,120	5.0	963	1001	n/a	n/a	n/a	n/a
Middle	4	160,133	17.5	1004	1004	27,514	7.2	960	998
	5	12,193	1.3	1024	1024	25,811	6.7	1006	1030
	6	193,853	21.2	1037	1037	178,598	48.0	1038	1038
	7	94,774	10.4	1043	1043	5,619	1.5	1045	1045
High	8	228,886	25.0	1075	1075	68,787	17.9	1077	1077

Tables B.2 and B.3 show the distribution of these smaller SEIFI and SEIFF groups across the IRSD deciles. As shown in Section 5, there is a negative relationship with between IRSD deciles and the lowest SEIFI and SEIFF groups. There is also a positive relationship for the highest SEIFI and SEIFF group. The middle SEIFI and SEIFF groups are fairly evenly distributed across each of the IRSD deciles.

B.2 Count of people in each IRSD decile by SEIFI score

IRSD decile	SEIFI groups 1	2	3	4	5	6	7	8	Total in decile
1	26,670	11,728	4,427	16,994	771	13,611	6,854	10,571	91,626
2	13,696	11,337	5,008	19,693	1,072	17,242	9,047	14,407	91,502
3	10,708	10,802	5,043	19,354	1,070	18,425	9,847	16,352	91,601
4	8,921	9,888	4,627	18,901	1,114	19,211	10,461	18,254	91,377
5	7,534	9,545	4,732	17,667	1,259	19,690	10,243	20,876	91,546
6	6,581	8,685	4,567	16,953	1,186	20,212	10,588	22,825	91,597
7	5,392	7,989	4,697	15,390	1,288	20,982	10,302	25,557	91,597
8	4,568	6,960	4,378	14,260	1,441	21,577	10,318	28,038	91,540
9	3,631	6,550	4,522	12,133	1,447	21,566	9,600	31,934	91,383
10	2,780	5,505	4,119	8,788	1,545	21,337	7,514	40,072	91,660
Total	90,481	88,989	46,120	160,133	12,193	193,853	94,774	228,886	915,429

B.3 Count of families in each IRSD decile by SEIFF score

IRSD decile	SEIFF groups 1	2	4	5	6	7	8	Total in decile
1	11,040	5,282	2,889	2,631	13,569	328	2,695	38,434
2	5,705	4,992	3,225	2,957	17,329	413	3,803	38,424
3	4,609	4,721	3,073	2,915	18,001	489	4,650	38,458
4	3,780	4,316	2,965	2,821	18,787	483	5,233	38,385
5	3,259	4,350	2,729	2,711	18,795	593	6,102	38,539
6	2,859	3,944	2,587	2,609	18,902	557	6,849	38,307
7	2,371	3,613	2,734	2,389	19,206	614	7,639	38,566
8	1,985	3,244	2,471	2,471	18,794	709	8,737	38,411
9	1,567	2,889	2,493	2,289	18,319	704	10,100	38,361
10	1,175	2,320	2,348	2,018	16,896	729	12,979	38,465
Total	38,350	39,671	27,514	25,811	178,598	5,619	68,787	384,350

FOR MORE INFORMATION . . .

INTERNET	**www.abs.gov.au** the ABS website 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 website 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 website, 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 statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS	www.abs.gov.au

2000001568309
ISBN 9780642483652

RRP $11.00