Introductory Guide

# Calculating a deprivation index using census data

**Paul Norman*** University of Leeds
**Laurie Berrie** University of Leeds
**Daniel J Exeter** University of Auckland

* Corresponding author. Email: p.d.norman@leeds.ac.uk. Address: School of Geography and Leeds Institute for Data Analytics, University of Leeds, Leeds LS2 9JT, United Kingdom

Paper received 30 October 2018; accepted 4 April 2019; published 27 May 2019

## Abstract

### Background

Deprivation indexes have widespread use in academic research and in local and national government applications. It is useful for people to understand their construction and to be able to calculate their own measures.

### Aims

We provide an overview of the background to area based deprivation measures. We detail and explain a series of steps taken to calculate a deprivation index for small areas in Australia.

### Data and methods

We use data from Australia's 2016 Census of Population and Housing for the SA2 level of geography. After defining the set of variables used as inputs, we emulate the steps taken to calculate other census based deprivation indexes.

### Results

The resulting scheme correlates closely with an official, but more sophisticated deprivation measure, suggesting that simple schemes have utility.

### Conclusions

There are choices to be made for input variables and for some of the detail of the calculations. Researchers can follow the steps we describe to develop their own measures.

### Key words

Area deprivation; composite measures; small areas; Census

## 1. Introduction

The use of area deprivation indexes is widespread across academic research and in local and national government applications (Allik et al. 2016). The Australian Bureau of Statistics produce Socio-Economic Indexes for Areas (SEIFA) using Census data, which includes the Index of Relative Socio-economic Disadvantage (IRSD). The latest version uses 2016 Census data.

Our aim here is to demonstrate a method of calculation to people, such as academic and local government researchers, who may be new to the production of composite measures. Following sections on generic aspects, this paper steps through the process of creating a deprivation index using census data for small areas (SA2s) in Australia for 2016. We take a similar approach to the calculation steps used in the long-established Townsend (Townsend et al. 1988) and Carstairs (Carstairs and Morris 1989) indexes. We compare the deprivation index calculated here with the SEIFA IRSD.

We provide data files so that researchers can reproduce the deprivation index and believe the method used here to be accessible to quantitative population geographers new to this kind of work who will then be able to construct their own composite measures. These measures might focus on particular dimensions of deprivation or be for other purposes for which a combination of indicator variables are applicable.

## 2. What is deprivation?

Deprivation is one of those things which we probably think we understand until we tried to define it! Townsend's (1987 p. 125) definition is commonly used: “Deprivation may be defined as a state of observable and demonstrable disadvantage relative to the local community or the wider society or nation to which an individual, family or group belongs. The idea has come to be applied to conditions (that is, physical, environmental and social states or circumstances) rather than resources and to specific and not only general circumstances, and therefore can be distinguished from the concept of poverty”. Defining poverty itself is far from easy (Paterson and Gregory 2019). From a Eurocentric perspective, deprived areas are places in which concentrations of people and households with the relevant characteristics exist.

## 3. What is a deprivation index?

A deprivation index provides a single figure estimate of the level of deprivation for each location within the study area. Since deprivation is both hard to define and cannot be directly measured, the index is a latent construct calculated using a combination of indicator variables believed to capture relevant dimensions. Having a single figure per area simplifies applications that use the index compared to using multiple variables. The deprivation index value for any one area is usually expressed relative to a larger study region, often the national level (Fu et al. 2015).

## 4. Which variables should be used as inputs?

As Noble et al. (2006 p. 172) point out, if a researcher is devising their own index, there should first be a clear theoretical foundation underpinning the notion of small area deprivation, which may then

be operationalised as an index. Deprivation indexes have their roots in census-based work in the UK, first by Holtermann (1975) and subsequently by Townsend (1987) and others. We cannot turn back the clock, but speculate that there may not have been the same motivation to construct deprivation indexes had the UK, unlike many other countries, included an income question in the census (Dorling 1999; Boyle and Dorling 2004). Thus, deprivation indexes in the UK use 'proxy' indicators of income and other dimensions of social and/or material hardship. Unemployment is ubiquitously included in schemes but there is variation in further variables used as inputs (see Senior 2002). In countries such as Australia, Canada and New Zealand, where income-related questions are asked, measures of individual and/or household income are components of their deprivation indexes.

The researcher needs to obtain variables which best represent particular aspects of deprivation to ensure the index is fit for purpose. The set of chosen variables should complement each other so that the index is 'greater than the sum of its parts'. The availability of potential variables for inclusion will be affected by decisions on the geographic unit used (below). Being based on data from a census, schemes are cross-sectional (though see the paper's concluding section). To an extent there will be overlap of the facets chosen to proxy aspects of deprivation and likely to be strong inter-relationships. Since the purpose is to identify multiple deprivation this collinearity may not be an issue. However, an index calculated in the manner described here would tend to have a few key variables. A more multi-dimensional approach would need a technique such as Principal Components Analysis (noted again below).

## 5. Which geography should be used?

Most deprivation indexes are devised for small areas (sub local government). As above, since the roots of schemes lie in the UK and the census has been the main source of data, then census geographies have tended to be used. Initially, within England and Wales, the most common geographical units used were the electoral wards (Mackenzie et al. 1998), while the Carstairs index, originally developed for Scotland, used Postal Sectors. These geographies were typically chosen in order to ensure that the count of people in each ward/sector was greater than 100, thus minimising the amount of missing / suppressed data.

People are often motivated to develop an index for the smallest possible areas, to aid regeneration targeting, for example. If the aim is to relate to geographical health variations, then a geography which represents people's day-to-day lives might be more suitable. Ideally, the researcher would choose the geography which best suited the relevant situation. In reality, researchers may well not have the freedom of choice of geography they might wish for, as the smallest unit for which the desired indicators are available may also dictate the level of analysis. It is important to note that all results will be affected by the specification of the zones which are used (Openshaw and Taylor 1981; Flowerdew et al. 2008).

## 6. A small area deprivation index for Australia

We now continue with the specifics involved with developing a small area deprivation index for Australia. Norman et al. (2016) calculated a measure of changing deprivation using data from the 2001 and 2011 Censuses in Australia. We will incorporate similar variables here for a cross-sectional

deprivation index for 2016. The geography will be the Statistical Areas Level 2 (SA2s). Whilst not the smallest census geography, the SA2s are local level areal units for which detailed Australian Bureau of Statistics data are released, including tables for geographic areas by a variety of person and household attributes. SA2s generally have populations between 3,000 and 25,000 persons with an average population of around 10,000 (ABS 2017).

### 6.1. Choose a set of indicator variables

Table 1 lists the variables we use. These have been chosen to represent a range of deprivation dimensions: unemployment (Haynes et al. 1996); low income (D'Ambrosio and Frick 2007); lack of good English (Bertotti et al. 2012); low educational achievement (Wilkinson and Pickett 2007) and; lone parents (Santana 2002). These are here as examples and each can be substituted or there can be additional variables should a researcher wish to use different indicators.

Table 1: Input variables from the 2016 Census in Australia (SA2)

| **2016 Census table number & title** | **Variable input to index** |
|---|---|
| G04 'Age by Sex' | Count of persons |
| G13 'Language Spoken at Home by Proficiency in Spoken English Language by Sex' | % Persons who do not speak English well |
| G16 'Highest Year of School Completed by Age by Sex' | % Persons not completing school to year 10 |
| G29 'Total Household Income (Weekly) by Household Composition' | % Households with income less than $1,500 (median income is $1,438) |
| G39 'Dwelling Structure by Household Composition and Family Composition' | % Dwellings with lone parent households |
| G43 'Labour Force Status by Age by Sex' | % Persons unemployed |

*Source*: Australian Bureau of Statistics
*Note*: The variables expressed as a percentage are Numerator / Denominator x 100.

The variables listed in Table 1 have been obtained from the table number stated. The numerators and denominators are sourced from the same tables so that appropriate percentages can be calculated. The inputs used for these are included in this article's online resources (noted at the end of the paper). The percentages represent variables whereby the higher the percentage, the greater the level of the deprivation. The polarity of the variables needs to be consistent.

### 6.2. Excluding areas with very small populations

In 2016 there were 2,301 SA2s in Australia but since some have very small populations, we only include those with 200 or more persons so the deprivation index developed is for 2,169 SA2s. Researchers could combine an area with a small population with an adjacent area so as not to lose data.

### 6.3. Transforming and standardising the input variables

Many social variables have skewed distributions and researchers may choose to transform input variables to (near) normal distributions. The Townsend index, for example, log transforms the

percentages of unemployment and overcrowding as these are positively skewed. The need for, and methods of transformations, are subject to debate (Gilthorpe 1995; Senior 2002). For simplicity, we do not transform variables here.

There is more of a consensus on the need to standardise variables. Underpinning the need for this step is that each of the input variables will have differences in what are inherently large and small values. For example, at the SA2 level, the percentage of persons not completing school to year 10 varies between 4.02% and 70.42% with a mean of 20.00, whereas the percentage of lone parents varies between 2.16% and 36.05% with a mean of 10.96. There is a need to have variables on a comparable scale.

The most common method (as used in both the Townsend and Carstairs schemes) is to standardise using Z-scores. The inputs to the Z-score calculation for the SA2 variables are the percentages of each variable. The resulting Z-scores for all variables have a mean of zero and a standard deviation of 1 so they are now comparable. Higher positive Z-scores are higher levels of the indicator and more negative values are lower levels. Z-scores are calculated using the formula:

Z-score = (Observation - Mean) / Standard Deviation

Alternatively, the 'range standardisation' calculation can be used (Lucy and Burns 2017). Range standardisation rescales observations for a variable to lie between 0 and 1.

## 6.4. Combining the standardised input variables

The aim is to have a single figure deprivation index and a regular approach is to sum the Z-scores of the input variables. A simple 'additive', unweighted approach is used in the Townsend and Carstairs indexes so that all variables contribute equally. The Jarman index (1983) applies weights to differentially change the impact of inputs when they are summed. In reality, the weights applied by Jarman more likely reflect GP workload rather than a variable's ability to better measure deprivation (Senior 2002).

For deprivation in Australia, we sum the five sets of Z-scores unweighted to thereby have a deprivation index for the SA2 geography. Once summed, in terms of interpretation, more positive index scores reflect higher levels of deprivation and more negative values are lower levels of deprivation. Since the indicators aim to capture deprivation, areas found to lack deprivation should not be described as relatively wealthy or affluent.

An alternative method of combining a set of indicator variables is to use Principal Components Analysis. The New Zealand Index of Deprivation for 2013 used nine census-derived variables of deprivation and the weights of the first principal components are used to combine the nine variables into an overall deprivation score (Atkinson et al. 2014). The SEIFA / IRSD in Australia also uses Principal Components Analysis (ABS 2018a).

## 6.5. Categorising into quantiles

Index scores themselves have great utility, but many applications use the scores categorised into quantiles. It is common to use quintiles or deciles whereby the distribution is divided into fifths or tenths. Often this is based on the number of areas or sometimes on the count of persons so that

there would be an equal number of areas or persons in each category. The impact of differently defined categorisations has had little coverage in the literature (compared with discussions of transformation and standardisation). The use of 'population weighted' quantiles has become common and is what we adopt here. The quantile cut-offs partition the deprivation distribution into categories with (near) equal numbers of people in each (Table 2).

Table 2: Counts of persons in each deprivation quantile

| **Deciles** | **Persons** | **Percent** | **Quintiles** | **Persons** | **Percent** |
|---|---|---|---|---|---|
| Q1 (least deprived) | 2,337,625 | 10.01 | Q1 (least deprived) | 4,671,384 | 20.01 |
| Q2 | 2,333,759 | 9.99 | | | |
| Q3 | 2,333,735 | 9.99 | Q2 | 4,668,398 | 19.99 |
| Q4 | 2,334,663 | 10.00 | | | |
| Q5 | 2,336,028 | 10.00 | Q3 | 4,672,799 | 20.01 |
| Q6 | 2,336,771 | 10.01 | | | |
| Q7 | 2,327,819 | 9.97 | Q4 | 4,660,545 | 19.96 |
| Q8 | 2,332,726 | 9.99 | | | |
| Q9 | 2,332,560 | 9.99 | Q5 (most deprived) | 4,677,180 | 20.03 |
| Q10 (most deprived) | 2,344,620 | 10.04 | | | |
| Total | 23,350,306 | 100.00 | Total | 23,350,306 | 100.00 |

Source: Based on Australian Bureau of Statistics data

## 7. Comparison of the deprivation index with an existing scheme

The Australian Bureau of Statistics produce Socio-Economic Indexes for Areas (SEIFA), which includes the Index of Relative Socio-economic Disadvantage (IRSD). The most recent version uses 2016 Census data. Note that there is a difference in polarity from the census measure we have developed here, so a low IRSD score indicates relatively greater disadvantage whereas a high score indicates a relative lack of disadvantage (ABS 2018b). The deprivation index developed in our analysis correlates -0.933 with the IRSD suggesting that the two schemes identify similar patterns of (non-) deprivation. A cross-tabulation of deciles from both schemes also demonstrates fair similarity in the categorisation of areas (Table 3).

Table 3: Cross-tabulation of the deprivation index and SEIFA IRSD quantiles

| | | **Deprivation Index** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Deciles | Q1 (least) | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 (most) | Total |
| **SEIFA IRSD** | Q1 (most) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 46 | 169 | 216 |
| | Q2 | 0 | 0 | 0 | 0 | 0 | 1 | 25 | 49 | 111 | 31 | 217 |
| | Q3 | 0 | 0 | 0 | 0 | 7 | 24 | 52 | 90 | 40 | 5 | 218 |
| | Q4 | 0 | 0 | 0 | 9 | 25 | 52 | 67 | 43 | 21 | 1 | 218 |
| | Q5 | 0 | 0 | 3 | 20 | 53 | 75 | 40 | 25 | 2 | 1 | 219 |
| | Q6 | 0 | 0 | 11 | 63 | 70 | 49 | 18 | 6 | 0 | 0 | 217 |
| | Q7 | 2 | 13 | 70 | 71 | 41 | 15 | 4 | 0 | 0 | 1 | 217 |
| | Q8 | 4 | 65 | 93 | 34 | 15 | 4 | 1 | 0 | 0 | 0 | 216 |
| | Q9 | 48 | 112 | 47 | 7 | 2 | 0 | 1 | 1 | 1 | 0 | 219 |
| | Q10 (least) | 177 | 34 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 212 |
| | Total | 231 | 224 | 225 | 204 | 213 | 220 | 208 | 215 | 221 | 208 | 2169 |

Source: Based on Australian Bureau of Statistics data

## 8. Conclusion

We have developed a deprivation index for the Australian SA2 geography using five indicators derived from the country's 2016 Census. We have taken similar steps to those used to calculate the Townsend and Carstairs indexes by calculating percentages of the variables, which are then standardised and summed (unweighted) to form a single figure deprivation index. We categorise the deprivation scores into deciles and quintiles. The similarity between an index produced in this relatively simple way with the more sophisticated SEIFA IRSD is consistent with previous findings in Australia (Norman et al. 2016).

Researchers wishing to develop their own indexes and make choices on geographic scale and input variables should be able to follow the same steps. It could be that an input to an existing scheme is regarded by others as an outcome of deprivation so there is a need for an index without that variable. The IMD for England includes a health indicator so there is risk of a self-fulfilling prophecy if relationships between deprivation and health are investigated (Adams and White 2004). Exeter et al. (2017) overcame this by providing a version of the IMD for New Zealand which removes a domain (i.e. IMD no Health) and redistributes the weights for the remaining six domains to reduce issues associated with data circularity. A composite index for a different application can be developed using relevant inputs. See, for example, Congdon (2004) on an index of social fragmentation and Lucy and Burns (2017) on an index of loneliness.

Two aspects will be relevant to people wanting to build on this work relating to alternative data sources or extending from cross-sectional measures to those which capture changing deprivation. A variety of deprivation schemes now use administrative data sources (e.g. Noble et al. 2006; Exeter et al. 2017). With a move away from reliance on census data, the same approach as we use here can have administrative data inputs. A scheme for England (Ajebon and Norman 2016) using a small number of administrative variables as inputs relates closely to both a census derived measure as well as the more sophisticated, multiple domain IMDs (Noble et al. 2006) now in common usage.

Census-based deprivation schemes are specific to the period of census data collection and administrative schemes for the year they are released (but may be based on inputs for different time points). All these are cross-sectional. However, areas change over time in their characteristics so increasingly researchers are interested in capturing this. Examples of methods and applications can be found in Norman (2010), Exeter et al. (2011; 2019), Norman et al. (2016), Pearce et al. (2016), Norman (2016), Green et al. (2017), Norman and Darlington-Pollock (2017) and Shackleton et al. (2018).

## 9. Resources

We provide an Excel file of the raw input variables plus commented files of SPSS syntax, Stata do and R script so that people can reproduce the process we have described above. A working knowledge of these packages is needed. The files can be accessed via: http://dx.doi.org/10.17632/k957gctr9d.1.

## Key messages

- This introductory guide explains the background concepts relating to area deprivation and to data and geography choices.
- The calculation of a deprivation index is stepped through and explained.
- Researchers are provided with a set of resources through which to reproduce the process.
- Researchers are encouraged to develop their own small area indexes.

### Acknowledgements

We are grateful for the feedback from two anonymous referees.

### References

Allik M, Brown D, Dundas R and Leyland A H (2016) Developing a new small-area measure of deprivation using 2001 and 2011 census data from Scotland. *Health & Place* 39*:* 122-130.

ABS (2017) Australian Statistical Geography Standard (ASGS). http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS). Accessed 17 May 2018.

ABS (2018a) Technical Paper Socio-Economic Indexes for Areas (SEIFA) 2016. Catalogue No. 2033.0.55.001. Canberra: ABS. http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf. Accessed 22 March 2019.

ABS (2018b) Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Catalogue No. 2033.0.55.001. Canberra: ABS. http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001. Accessed 22 March 2019.

Adams J and White M (2006) Removing the health domain from the Index of Multiple Deprivation 2004 - effect on measured inequalities in census measures of health. *Journal of Public Health* 28(4): 379-383.

Ajebon M and Norman P (2016) Beyond the census: a spatial analysis of health and deprivation in England. *GeoJournal* 81(3): 395-410.

Atkinson J, Salmond C and Crampton P (2014) NZDep2013 Index of Deprivation. University of Otago (Wellington) https://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html. Accessed 27 September 2018.

Bertotti M, Adams-Eaton F, Sheridan K and Renton A (2012) Key barriers to community cohesion: views from residents of 20 London deprived neighbourhoods *GeoJournal* 77(2): 223-234.

Boyle P and Dorling D (2004) The 2001 UK census: remarkable resource or bygone legacy of the 'pencil and paper era'? *Area* 36(2): 101-110.

Carstairs V and Morris R (1989) Deprivation: explaining differences in mortality between Scotland, England and Wales. *British Medical Journal* 299(6704): 886-889.

Congdon P (2004) Commentary: Contextual effects: index construction and technique. *International Journal of Epidemiology* 33(4): 741-742.

D'Ambrosio C and Frick J R (2007). Income satisfaction and relative deprivation: An empirical link. *Social Indicators Research* 81(3): 497-519.

Dorling D (1999) Who's afraid of income inequality? *Environment & Planning A* 31(4): 571-574.

Exeter D J, Boyle P J and Norman P (2011) Deprivation (im)mobility and cause-specific premature mortality in Scotland. *Social Science & Medicine* 72(3): 389-397.

Exeter D J, Zhao J, Crengle S, Lee A and Browne M (2017) The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand. *PLoS ONE* 12(8): e0181260.

Exeter D J, Feng Z, Zhao J, Cavadino A and Norman P (2019) The geographic harmonisation of Scotland's small area census data, 1981 to 2011. *Health & Place* 57: 22-26.

Flowerdew R, Manley D J and Sabel S E (2008) Neighbourhood effects on health: does it matter where you draw the boundaries? *Social Science & Medicine* 66(5): 1241-1255.

Fu M, Exeter D J and Anderson A (2015) "So, is that your 'relative' or mine?" A political-ecological critique of census-based area deprivation indices. *Social Science & Medicine* 142: 27-36.

Gilthorpe M (1995) The importance of normalisation in the construction of deprivation indices. *Journal of Epidemiology and Community Health* 49(S2): S45-S50.

Green M A, Arcaya M and Subramanian S V (2017) Using internal migration to estimate the causal effect of neighbourhood socioeconomic context on health: a longitudinal analysis (England, 1995-2008). *Annals of the Association of American Geographers* 107(6): 1266-1278.

Haynes R, Gale S, Lovett A and Bentham G (1996) Unemployment rate as an updatable health needs indicator for small areas. *Journal of Public Health* 18(1): 27-32.

Holtermann S (1975) Areas of urban deprivation in Great Britain: an analysis of 1971 Census data. *Social Trends* 6: 33-45.

Jarman B (1983) Identification of underprivileged areas. *British Medical Journal* 286: 1705-1709.

Lucy L and Burns L (2017) Devising a composite index to analyze and model loneliness and related health risks in the United Kingdom. *Gerontology and Geriatric Medicine* 3: 1-10.

Mackenzie I F, Nelder R, Maconachie M and Radford G (1998) My ward is more deprived than yours. *Journal of Public Health Medicine* 20(2): 186-190.

Noble M, Wright G, Smith G and Dibben C (2006) Measuring multiple deprivation at the small-area level. *Environment and Planning A* 38(1): 169-185.

Norman P (2010) Identifying change over time in small area socio-economic deprivation. *Applied Spatial Analysis and Policy* 3(2-3): 107-138.

Norman P (2016) The changing geography of deprivation in Britain, 1971 to 2011 and beyond. In Champion T and Falkingham J (eds.) *Population Change in the United Kingdom*. London: Rowman and Littlefield; 193-214.

Norman P and Darlington-Pollock F (2017) The changing geography of deprivation in Great Britain: exploiting small area census data, 1971 to 2011. In Stillwell J (ed.) *Routledge Handbook of Census Resources, Methods and Applications*. Abingdon, Oxon: Routledge; 404-420.

Norman P, Charles-Edwards E and Wilson T (2016) Relationships between population change, deprivation change and health change at small area level: Australia 2001-2011. In Wilson T, Charles-Edwards E and Bell M (eds.) *Demography for Planning and Policy: Australian Case Studies*. Cham: Springer; 197-214.

Openshaw S and Taylor P J (1981) The modifiable areal unit problem. In Wrigley N and Bennett R J (eds.) *Quantitative Geography: A British View*. London: Routledge and Kegan Paul; 335-350.

Paterson L L and Gregory I N (2019) Defining and measuring poverty. In *Representations of Poverty and Place*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-93503-4_1

Pearce J, Shortt N, Rind E and Mitchell R (2016) Life course, green space and health: incorporating place into life course epidemiology. *International Journal of Environmental Research in Public Health* 13(3): 331.

Santana P (2002) Poverty, social exclusion and health in Portugal. *Social Science & Medicine* 55(1): 33-45.

Senior M L (2002) Deprivation indicators. In Rees P, Martin D and Williamson P (eds.) *The Census Data System*. Chichester: John Wiley; 123-137.

Shackleton N, Darlington-Pollock F, Norman P, Jackson R and Exeter D (2018) Longitudinal deprivation trajectories and risk of cardiovascular disease in New Zealand. *Health and Place* 53: 34-42.

Townsend P (1987) Deprivation. *Journal of Social Policy* 16(2): 125-146.

Townsend P, Phillimore P and Beattie A (1988) *Health and Deprivation: Inequality and the North*. Andover: Croom Helm.

Wilkinson R G and Pickett K E (2007). The problems of relative deprivation: why some societies do better than others. *Social Science & Medicine* 65(9): 1965-1978.