Chapter 3 Methodology

Section 3.1: Use of the 2001 Census

As indicated, each PIMD is based on the 2001 Census. The Census contains a wealth of valuable information on personal and household circumstances and many direct measures of deprivation.

Every person present in South Africa on Census night 9-10 October 2001 should have been enumerated in the 2001 Census. Thus the Census provides useful information on the whole population in one dataset. Imputation was carried out on the full Census by Stats SA to allocate values for unavailable, unknown, incorrect or inconsistent responses. A combination of 'logical' imputation and 'hot deck' imputation was used when inconsistencies were found in the data[11].

Stats SA has made available a 10% sample of the 2001 Census. The 10% sample is a useful and easy to use dataset for testing different indicators and combinations of indicators to be used in the Provincial Indices of Multiple Deprivation. Using the 10% sample it was only possible to produce an index at municipality level as the data is only robust to this level and ward codes are not supplied. However, once the final set of indicators, domains and combination techniques had been decided on, it was possible to run the entire code (with appropriate amendments) on the full dataset to produce each PIMD at ward level.

Section 3.2: Creating domain indices

Dealing with small numbers

To improve the reliability of a score which is based on small numbers, the shrinkage estimation technique can be applied. The effect of shrinkage is to move the score for a small area towards the average score of a larger area for a particular indicator. For example, where wards are the small area geography, the ward level scores would be moved towards the average score for the municipality in which the ward is located. The extent of movement depends on both the reliability of the indicator and the heterogeneity of the larger area. If scores are robust, the movement is negligible as the amount of shrinkage is related to the standard error. The shrinkage technique does not mean that the score necessarily becomes smaller (i.e. less deprived). Where wards do move this may be in the direction of more deprivation if the 'unreliable' score shows less deprivation than the municipality mean. For further details about the shrinkage technique, see the **Technical Report**.

[11] Further details on the imputation techniques used, and also the Census in general, are available from Stats SA.

The impact of shrinkage was tested on all domains[12], but it was found that there was very little movement in the scores, and so for transparency of method, the 'unshrunk' scores were used for all indicators, other than the Years of Potential Life Lost indicator in the Health Deprivation Domain where the 'shrunk' score was used.

Combining indicators into domain indices

For each domain of deprivation (Income, Employment, etc) the aim is to obtain a single summary measure whose interpretation is straightforward in that it is, if possible, expressed in meaningful units (e.g. proportions of people or of households experiencing that form of deprivation). Apart from the Health Deprivation Domain, all of the other domains were created as simple rates. This avoided the key issue of weighting indicators which is necessary when combining indicators into a single measure. Because the domain scores are rates they are easy to interpret (i.e. X% of people in the ward of the relevant age are experiencing this type of deprivation). As discussed in **Section 2.4**, the Health Deprivation Domain is more complex as it had to be age standardised and the technique of shrinkage estimation was applied to ensure robustness.

There is no double counting of individuals within a domain. An individual may be captured in more than one domain but this is not double counting: it is simply identifying that they are deprived in more than one way.

After combining the indicators, District Management Areas[13] (DMAs), and fragments of split wards where the population was less than 100, were omitted from each PIMD, leaving the following number of wards in each province:

Western Cape	332
Eastern Cape	604
Northern Cape	153
Free State	291
KwaZulu-Natal	750
North West Province	375
Gauteng	420
Mpumalanga	361
Limpopo	487

Five domain indices were created for each province which were then combined into an overall PIMD.

[12] This testing was undertaken at ward level for the whole country. See the **Technical Report**

[13] District Management Areas are areas such as game reserves and mining complexes with small populations with special characteristics. They produce anomalous results and are customarily excluded by Stats SA from small area analyses.

Section 3.3: Combining domain indices into an index of multiple deprivation

Standardisation and transformation

Domains are conceived as independent domains of deprivation, each with their own contribution to multiple deprivation. The strength of this contribution should vary between domains depending on their relative importance. Once the domains had been constructed, it was necessary to combine them into an overall index for each province. In order to do this the domain indices were standardised by ranking. They were then transformed to an exponential distribution.

The exponential distribution was selected for the following reasons. First, it transforms each domain so that they each have a common distribution, the same range and identical maximum/minimum value, so that when the domains are combined into a single index of multiple deprivation the (equal) weighting is explicit; that is there is no implicit weighting as a result of the underlying distributions of the data. Second, it is not affected by the size of the ward's population. Third, it effectively spreads out the part of the distribution in which there is most interest; that is the most deprived wards in each domain.

Each transformed domain has a range of 0 to 100, with a score of 100 for the most deprived ward. The exponential transformation that was selected for standardising the domains in the ward level PIMD stretches out the most deprived 25% of wards in each province. The chosen exponential distribution is one of an infinite number of possible distributions. Two other exponentials were explored: stretching out the most deprived 10% of wards (used in UK Indices) and stretching out the most deprived 30% of wards. When transformed scores from different domains are combined by averaging them, the skewness of the distribution reduces the extent to which deprivation on one domain can be cancelled by lack of deprivation on another. For example, if the transformed scores on two domains are averaged with equal weights, a (hypothetical) ward that scored 100 on one domain and 0 on the other would have a combined score of 50 and would thus be ranked at the 75th percentile. (Averaging the untransformed ranks, or after transformation to a normal distribution, would result in such a ward being ranked instead at the 50th percentile: the high deprivation in one domain would have been fully cancelled by the low deprivation in the other). Thus the extent to which deprivation in some domains can be cancelled by lack of deprivation in others is, by design, reduced. The exponential transformation procedure is set out in more detail in the **Technical Report**.

There are a number of other ways in which domain scores could have been standardised/transformed prior to combination. Examples include 'z' scores and the 'signed chi square' technique. However each has major drawbacks. The former leads to unpredictable *implicit* weighting where there are significant outliers at either end of the distribution; the latter introduces size of population into the equation in an unpredictable way (for a discussion see Noble *et al*., 2000a: 53-56). In the case of the UK work, ranking and then transforming the ranks to an exponential transformation distribution proved

most satisfactory (Noble *et al*., 2000a). For this reason the technique was used with modification in the South African situation. In UK work the most deprived 10% of the distribution was 'spread out', whereas in South Africa with much higher levels of deprivation it was thought that it was appropriate to spread out the most deprived quarter of the distribution. Some sensitivity testing on different standardisation/transformation techniques and different exponential distributions was, however, undertaken and the analysis appears in the **Technical Report**.

As work in this area develops it is recommended that further work be undertaken on methods of combination of the domain indices to construct the overall Indices of Multiple Deprivation.

Weighting

An important issue in constructing an overall index of multiple deprivation is the question of what 'explicit weight' should be attached to the various components. The weight is the measure of importance that is attached to each component in the overall composite measure. How can one attach weights to the various aspects of deprivation? That is, how can one determine which aspects are more important than others?

There are at least five possible approaches to weighting:

a) driven by theoretical considerations;
b) empirically driven;
c) determined by policy relevance;
d) determined by consensus; and
e) entirely arbitrary.

In the theoretical approach, account is taken of the available research evidence which informs the theoretical model of multiple deprivation and weights are selected which reflect this theory.

There are two sorts of empirical approaches that might be applicable. First a commissioned survey or re-analysis of an existing survey might generate weights. Second one might apply a technique such as factor analysis to extract some latent 'factor' called 'multiple deprivation', assuming that is, that the analysis permitted a single factor solution (see Senior, 2002).

Alternatively, the individual domain scores could be released and weighted for combination in accordance with and proportional to the focus of particular policy initiatives or weighted in accordance with public expenditure on particular areas of policy.

Another approach would be for policy makers and other 'customers' or experts to simply be consulted for their views and the results examined for consensus.

Finally, simply choosing weights without reference to the above or even selecting equal weights in the absence of empirical evidence would come into the category of ‘entirely arbitrary’. Weighting always takes place when elements are combined together. Thus if the domains are summed together to create an index of multiple deprivation this means they are given equal weight. It would be incorrect to assume that items can be combined without weighting.

For each PIMD, equal weights were assigned to the exponentially transformed domains in the absence of evidence suggesting differential weights should be used.

Appendix 2 summarises the components of each PIMD in diagrammatic form.