EQUINET
methods toolkit

Deprivation & Resource Allocation:

Methods for small area research

Health Economics Unit,
University of Cape Town
and Centre for Health Policy,
University of the Witwatersrand
for the Regional Network for
Equity in Health in Souhern Africa
(EQUINET)

SEPTEMBER 2003

With support from IDRC (Canada)

Deprivation and Resource Allocation: Methods for small area research

EQUINET methods toolkit

Health Economics Unit, University of Cape Town and
Centre for Health Policy, University of the Witwatersrand
for the Regional Network for Equity in Health in Southern Africa
(EQUINET)

September 2003
with support from IDRC (Canada)

Undertaking small area research to explore deprivation and resource allocation issues: Key considerations

Health Economics Unit, University of Cape Town
Centre for Health Policy, University of the Witwatersrand

Introduction

There is a growing interest internationally in undertaking studies which consider health disparities between small geographic areas, both as a tool to understand better the determinants of health inequalities and to explore appropriate policy responses. One of the potential policy responses is to consider ways in which government resources can be allocated to redress health inequalities. Small area analyses can be used to identify locations with the greatest health need, and then to give preference in the allocation of health care resources to these areas. A study was recently undertaken to consider these issues in the South African context.[1] This study focused on estimating deprivation in small areas and analysed the distribution of deprivation between these areas. The relationship between deprivation and ill health was also explored. Finally, this study considered how deprivation indicators could be taken into consideration when determining the allocation of public sector resources.

The purpose of this document is to provide a guide to the main steps in completing a small area analysis into deprivation and resource allocation using lessons learnt in the South African study. It is hoped that this will facilitate similar research being undertaken in other countries in the SADC region. This document should be read in conjunction with the research report1, which provides greater detail on certain conceptual issues as well as insights into the interpretation and analysis of the data.

[1] McIntyre D, Muirhead D, Gilson L, Govender V, Mbatsha S, Goudge J, Wadee H, Ntutela P (2001).
Geographic patterns of deprivation and health inequities in South Africa: Informing public resource allocation strategies.
Harare: Equinet.

Steps in conducting small area research of deprivation and resource allocation

step 1

DETERMINING HOW TO VIEW DEPRIVATION

The concept of deprivation relates to relative social and material disadvantage. Thus, it refers to the material and social conditions that are experienced by individuals and households, where these conditions are inadequate relative to what is usually available or experienced in society. Deprivation is a broader concept than poverty, which traditionally has been defined as insufficiency of income.

A large number of studies have attempted to measure deprivation, most of which have been conducted in the United Kingdom. More recently, a number of studies have been undertaken in middle-income countries. Despite the growing body of literature, there exists no definitive method of measuring deprivation. However, common to all these measures is the combination of a number of socio-economic and demographic variables into a composite index of deprivation. The key factors differentiating the indices from each other are the selection of their component variables, and whether the variables are weighted equally or differentially to form the composite deprivation index. Table 1 highlights the kinds of variables most frequently included in deprivation indices in different country contexts. While there are some similarities between the variables used in the different country contexts, there is a much greater emphasis on lack of access to basic facilities (e.g. potable water, sanitation, safe energy sources) in middle-income countries.

Table 1: Variables frequently included in deprivation indices in different contexts

Variables frequently included in deprivation indices in high-income countries	Variables frequently included in deprivation indices in middle-income countries
Unskilled worker/Low social class	Illiteracy/low educational attainment
Unemployment	Lack of access to running water
Overcrowding in housing	Lack of access to electricity
Socio-economic group facilities	Lack of access to sanitation/sewerage
Child under the age of 5	Low quality housing
Pensioner living alone	Overcrowding in housing
Belonging to a minority ethnic group	Low income levels
Changed house/address in past year (Mobility)	Unemployment
Don't own a car	Extent of debt
Single parent	Lack of assets/durable household goods
Living in rented accommodation/ don't own a house	Age (children and the elderly may be more deprived)
Lack of amenities (shower & inside toilet)	Gender (women may be more deprived)
Lack of educational qualifications	Geographic area (rural dwellers)

The choice of which variables to include in an analysis of deprivation will vary from country to country.
The selection of variables will be strongly influenced by what data are available.

step 2

EXPLORING AVAILABLE DATABASES

Types of data needed

There are broadly four categories of variables that are valuable for this type of analysis. These are:

- Demographic variables (e.g. age, gender);
- Socio-economic variables that are specific to the individual (e.g. educational status);
- Socio-economic variables that apply to a household (e.g. type of sanitation, overcrowding); and
- Health status indicators.

Potential sources of data

The national census database often has a variety of demographic and socio-economic variables. It also has the benefit of including the majority of the population, if not the entire population, so the numbers involved are large. This enables you to undertake a wide range of statistical analyses at a small area level. However, a national census usually represents income in terms of categories (e.g. $0-$500; $501-$1,000) which limits the use of income data. In addition, a census very rarely includes health status indicators.

In most low- and middle-income countries, there are some household surveys that include the four categories of variables mentioned above. This includes surveys such as the Living Standards Measurement Survey (LSMS). Unfortunately, one of the household surveys undertaken in a large number of low- and middle-income countries, the Demographic and Health Survey (DHS), has very limited socio-economic data which restricts its usefulness in this type of work. Some countries may have routine national household surveys that may include relevant variables. The major drawback with household survey databases is that they often have a relatively small sample size. This limits the statistical analyses that can be done and conclusions that can be proposed.

In the absence of health status data from any of the above types of surveys, you may have to rely on vital statistics (i.e. data from registration of births and deaths). However, the use of these data can have some specific problems. In low- and middle-income countries, there is a tendency for a lower proportion of the deaths occurring in rural areas to be registered than in urban areas. When indicators such as mortality rates are then calculated, a false picture can be created that residents in rural areas have better health status than urban. In addition, it may be difficult to get this data according to the small areas you wish to use.

In summary each type of database has advantages and disadvantages. These need to be assessed within the specific country context and the level of analysis you wish to perform.

Selecting appropriate databases and combining them

In selecting potential databases for use, three guiding principles should be borne in mind:

- You should attempt to get as wide a range of variables as possible – although you may have ideas on what variables may be important in measuring deprivation, one should not rely too heavily on preconceptions as this may limit the analysis.
- The database(s) should have a sufficiently large sample size at a small area level to enable statistical analysis.
- The data should be available in a way that allows analysis using different sizes of geographic area. This will allow one to use a smaller geographic area if it is found that the geographic area initially selected does not have a sufficiently homogenous population (see later discussion of homogeneity analysis). Surveys used in small area analyses will generally have codes, or at least names, that represent each small geographic area (e.g. enumerator area, ward, municipality).

Though the ideal may be to have access to the full range of these variables in one database, this is rarely possible. However, you may have access to a number of different databases that, when used together, could provide the range of information necessary. The databases can be combined at the small area level (see Box 1).

Box 1: Illustration of how to combine datasets

Census:

Information for each small area on:

- Demographic (age and gender)
- Household-level socioeconomic characteristics (e.g. housing type, overcrowding, access to water and sanitation, etc.)

Living Standards Measurement Survey:

Information for each small area on:

- Household-level socioeconomic characteristics (e.g. average per capita household income)
- Individual-level socioeconomic characteristics (e.g. education)

Vital Statistics:

Information for each small area on:

- Mortality (e.g. Infant Mortality Rate, Standardised mortality rates)

Combined dataset:

Information for each small area from census, LSMS and vital statistics databases (combined together via small area, i.e. the small area code or name* used to link data from 3 separate databases)

* Note: Remember to check that the small areas used are consistent between datasets. Even if the same codes or names are used, the area boundaries that this represents could have been changed between the surveys.

See Appendix B (Section 1. COMBINING DATASETS) for statistical commands in SPSS and STATA

step 3

SELECTING THE SMALL AREA

A range of different sizes of areas (both in terms of physical and population size) have been used in what are classified as "small area" studies. There are problems both with "going too small" and of "not going small enough". The main concern with relatively large areas is that they are less likely to contain a homogenous population. This may mean that while an area on average has relatively good socio-economic indicators, it may contain pockets with very poor socio-economic status. The key problem with extremely small areas is that there may be inadequate numbers in certain variables to allow for statistical modelling.

While there is no agreement on exactly what size a small area should be, the international literature provides some useful guidelines on how to determine an appropriate size within specific country and study contexts. In particular, the following factors should be taken into consideration:

- Availability of information – one needs to consider the usual level of disaggregation of data in alternative databases that contain variables needed for analysis in the study (e.g. do the various databases provide information in terms of enumeration areas or only at a higher level of aggregation such as a ward or municipality);
- Physical and population size – it is also important to assess whether the preferred small area categories will yield sufficient numbers to permit statistical analysis; and
- Homogeneity – evaluating whether there is wide variation in the variables being measured within the preferred small area (e.g. if one is exploring differences in average household income between small areas, there should not be a wide range of household incomes within each small area).

step 4

BASIC PRINCIPLES FOR COMPILING A DEPRIVATION INDEX

A deprivation index, particularly the variables selected to form part of the index, should be specific to the country in which it is be used. Useful principles that can be used to guide the development of a country-specific deprivation index include:

- Indices should follow from the policy goals – It is important to clearly state the policy goals of the study, and to ensure that index development is based on these goals.
- The variables included in the index should be additive – The concept ‘additive’ means that if an individual ranks poorly with regard to two or more variables included in an index, that individual is more likely to be deprived than an individual belonging to only one of the categories. For example, if an index is constructed from 2 variables, which are ‘elderly person (over the age of 60)’ and ‘living alone’, then elderly people living alone are likely to be more deprived than either elderly people who do not live alone or people who live alone but are not over the age of 60.[2]
- Different weights should be assigned to each variable in the index – If individual variables are not weighted (i.e. each variable’s value is simply added to the value for every other variable to form a composite index), there is an implicit, and often false, assumption that each variable or indicator contributes to deprivation to the same extent as every other variable in the index. This means that individuals displaying any one characteristic reflected in the index are just as likely to experience deprivation as individuals or households displaying any other characteristic. Assigning weights to each variable makes explicit the relative importance of different variables/indicators in driving deprivation.

A statistical technique (called principal component analysis - PCA) exists that ensures that an index includes variables that are additive and assigns different weights to each variable. There is a growing consensus in the international literature that PCA is the preferred technique for developing a country-specific deprivation index. For this reason, PCA is described in some detail in a later section. However, it should be noted that the South African study found that much simpler indices were almost as effective in identifying small areas with high levels of deprivation as the more complex index developed using PCA techniques. Simpler indices may have a smaller number of variables that are not weighted or may be a single variable that has been shown to be an important indicator of deprivation. These have some distinct advantages over a more complex composite index, particularly in their ease of calculation aiding ongoing monitoring. The degree of information that may be lost taking this approach, though, should be determined prior to settling on a single variable indicator of deprivation.

[2] Two useful references on this topic are Gordon, 1995 and Folwell, 1995 – see Appendix A

step 5

UNDERTAKING A PRINCIPLE COMPONENT ANALYSIS

Introduction

At this stage, you may have large numbers of variables that you are considering including in a deprivation index. Factor analysis is a technique that can do two important things:

- Firstly, it can guide you as to the variables to include in an index of deprivation
- Secondly, it can produce weights for each of your variables from the data itself

Factor analysis combines individual variables that are highly correlated with each other into subsets, each subset being relatively independent of (uncorrelated with) the others.

Principal component analysis (PCA) is a particular type of factor analysis. It identifies which variables interact with each other and identifies the 'component' (combination of variables) that explains the interaction between these variables most comprehensively. Stated differently, PCA identifies the most important relationships between variables. More details of factor analysis and principal component analysis can be found in the literature referred to in the annotated bibliography (see Appendix A).

Selecting and preparing the variables to include in the PCA

As indicated previously, one should select a relatively wide range of possible variables initially, depending on data availability. The selection of variables to include in the initial analysis can be guided by international experience of the variables that are likely to be of greatest relevance (see Step 1) and by considering the relevance to the country in which the analysis is being conducted. For example, it may be important to review recent policy documents and/or to interview stakeholders about what characteristics or variables are most likely to contribute to deprivation in that country. One could include just demographic and socio-economic variables, in order to explore general deprivation. Alternatively, if one wanted to explore health-related deprivation, one could include a health-related variable (e.g. Standardised Mortality Ratios) along with the demographic and socio-economic variables.

Once the variables have been chosen, it is necessary to ensure that the data for these variables is translated into a useable format. For example, if the variable 'unemployment' were selected, data would most appropriately be presented as the percentage of the economically active population who are unemployed in each small area. If one simply uses the total number of people who are unemployed in each small area, an area that has a large number of unemployed people may be considered relatively deprived even if it has a large population size and hence a relatively low unemployment rate. Essentially, one needs to express the specific variable in relation to the underlying population of relevance to that variable.

A further aspect of preparing the data relates to transforming data for variables that are not symmetrically distributed. If the underlying distribution of any of the variables you are using is skewed, false relationships between variables can be produced and therefore the factor analysis biased. The distribution of each variable to be used should be looked at through a histogram or by producing statistics of skewness and kurtosis prior to undertaking the factor analysis. Any variables showing skewed distributions should be transformed prior to placing them in the factor analysis procedure. The most common types of transformations are the square root, reciprocal and logarithm with their corrective power increasing respectively. After applying the transformation, the variables' distribution should again be checked until the skewness is reduced to as close to zero as possible.[3] Once these steps have been undertaken, it is necessary to make the final selection of variables to include in the PCA. There are two elements to this process, namely ensuring that all variables are highly correlated with each other and ensuring that variables are additive. To investigate the correlation in data, all socio-economic and demographic variables selected above should be included in a bivariate Spearman rank correlation analysis. Variables that show a high correlation with all other socio-economic and demographic variables (defined as significant at 1% level) may be included in the PCA. The final stage is then to ensure that each of the variables can be considered to be additive (see Step 4). It is particularly important to assess whether any of the variables may lead to duplication of 'double-counting'. For example, if the variables 'proportion of household heads who are unemployed' and 'overall proportion of unemployed' in the small area are highly correlated, including both variables in the PCA may lead to a duplication of the effect of unemployment.

Finalising the selection of small areas for analysis

As indicated previously, the type of small area (e.g. enumeration area, ward or municipality) selected is also heavily influenced by data availability (see Step 3). However, it is important to ensure that the preferred small areas do have a relatively homogenous population. The extent of homogeneity can be assessed by calculating the coefficient of variation (= standard deviation ÷ mean) for key variables within each small area. If this analysis shows that there is considerable variation in key variables within the small areas, it may be necessary to select areas that are even smaller than initially anticipated.

Undertaking the PCA

Most statistical packages have a function for undertaking different types of factor analysis, including PCA.

See Appendix B (Section 2. UNDERTAKING THE PCA) for statistical commands in SPSS and STATA

As indicated previously, the PCA will produce a series of components (i.e. subsets of highly correlated variables), with each component or subset being relatively independent of, or uncorrelated with, the other components. When undertaking the PCA, it should include a 'varimax rotation'. Varimax rotation simplifies the results in that it ensures that the components extracted reduce the number of variables of importance within each

[3] See Appendix A for Gilthorpe (1995) reference on normalisation.

component providing the simplest solution. An example of a PCA output is shown in Table 2 below.

Table 2: Components arising from analysis of general deprivation

The PCA output has 3 components. The variables of importance in each component are the unshaded ones in the respective columns. The first component indicates that a large number of variables are contributing to that particular aspect of deprivation. The second component indicates that 'race', unemployment and disability seem to interact to contribute to certain types of deprivation, while the third component indicates that gender and being elderly interact to contribute to other aspects of deprivation. The 'loadings' on each variable (i.e. the numbers next to each variable in the components) indicate the relative importance of each variable in contributing to deprivation (e.g. the proportion of households that have a phone is the most important variable contributing to deprivation in the above example). The first component is the one that should form the basis of the deprivation index as it explains the largest proportion of the variance in correlations between small area characteristics of all of the three components.

	1	2	3
FEMALE	0.556	0.262	0.599
CHILD	0.872	0.075	0.077
RURAL	0.876	0.055	0.020
BLACK	0.488	0.744	-0.281
NO SCHOOL	0.800	0.163	0.111
UNE MP	0.529	0.672	0.097
DISAB	-0.006	0.915	0.100
HOUSE	0.721	0.404	-0.238
WATER	0.877	0.344	0.043
REFUSE	0.803	0.151	0.021
PHONE	0.892	0.223	0.063
LIGHT	0.820	0.316	0.010
FEHMD	0.782	0.405	0.235
ELDERLY	-0.033	-0.067	0.914
% TOTAL VARIANCE EXPLAINED	50.056	18.305	10.208

In order to calculate the final deprivation index, factor scores must be produced using a statistical package when completing the PCA (see Appendix B). This will produce component score coefficients which will act as weights for each variable in the final index. It should be noted that these coefficients are not the same as the 'loadings' found in the PCA (compare the coefficients in the equation below with the 'loadings' in Table 2 for each variable). The standard scores (z-scores) should also be calculated for each variable that contributed significantly to the component. The index is calculated by summing the weighted variables, and can be represented (for the PCA output presented in the Table above) in simplified terms as follows (where each of the variable labels represents the z-score of that variable):

Deprivation index = (0.028 x FEMALE) + (0.181 x CHILD) + (0.190 x RURAL)
+ (0.141 x NOSCHOOL) + (0.040 x UNEMP) + (0.091 x HOUSE)
+ (0.124 x WATER) + (0.151 x REFUSE) + (0.152 x PHONE)
+ (0.117 x LIGHT) + (0.072 x FEMHD)

Once the deprivation index value has been calculated for each small area, various analyses can be undertaken, as discussed in the next section.

step 6

EXPLORING THE DISTRIBUTION OF DEPRIVATION

As a first step, it may be useful to rank the small areas according to their deprivation index value. This will assist in identifying those areas that experience the highest levels of deprivation. One can also present the information in terms of quintiles of deprivation index values (i.e. allocate small areas to the most deprived 20% of small areas, the next most deprived 20% of small areas, etc.). This can then be used to show the distribution of deprivation across the country, either in the form of bar charts (e.g. percentage of most deprived small areas/districts in each province/region) or in the form of maps (e.g. colour coding small areas/districts according to the 5 quintiles). The use of maps is particularly effective in drawing politicians' attention to the distribution of deprivation in the country. It is easy to identify the most deprived areas in this format, and politicians may then be more likely to support initiatives to secure and target additional resources to relatively deprived areas.

It may also be worthwhile comparing the distribution of deprivation with the distribution of health-related indicators (such as mortality). Although numerous studies in high-income countries have found a significant relationship between deprivation and ill health, this has not been explored in much detail in low- and middle-income countries. Such an analysis may once again be powerful in persuading politicians about resource allocation priorities as one is highlighting that certain communities are not only deprived in relation to a range of socio-economic and other factors influencing relative disadvantage, but also in relation to their health status.

Finally, it may be valuable to undertake a comparison of the distribution of deprivation and health indicators with that of health service indicators. In many countries it has been found that the areas that are most deprived and bear the greatest burden of ill-health are also the areas that have the least access to health services / have the lowest levels of health care resources. This provides yet another persuasive element to arguing for equity to be a driving force in health care resource allocation decision-making.

step 7

INCORPORATING THE INDEX INTO A RESOURCE ALLOCATION FORMULA

Introduction to needs-based resource allocation formulae

In an effort to promote geographic equity, an increasing number of countries are basing their decisions about the allocation of public sector resources between geographic areas (e.g. provinces, regions, districts) on formulae which include measures of relative need for health care within particular geographic areas. The size of the population in each geographic area is the primary indicator of need for health services used in such formulae. Population size can then be weighted by a range of other indicators of the relative need for health care, such as:

- The demographic composition in each area, to account for the different health care needs of different age and gender categories;
- Mortality levels in each area, such as standardised mortality ratios (SMRs), to account for different levels of ill-health between geographic areas;
- The level of deprivation in each area as this may not only influence the level of ill health in an area but also indicate the extent to which communities are able to pay for the costs of health care and hence the level of dependence on publicly financed health services.

Thus, the deprivation index calculated through the methods described above can be used in a resource allocation formula to guide health care decision-makers.

Calculating a resource allocation formula that includes a deprivation index

The first step is to 'normalise' the deprivation index. The index values will be negative for some geographic areas (those that are least deprived) and positive for other areas (those that are most deprived). Normalising the index results in the least deprived area having a value of 1 and all other areas being expressed in relation to the least deprived area's value. In essence, one is 'shifting the axis' across so that the lowest value is 1. Table 3 below indicates the deprivation index value for provinces in one country. It indicates that the least deprived province (Province C) has an index value of –1.18893. In order to normalise the index values, one needs to add 2.18893 to the index value in each province, thereby making Province C's normalised value 1 with all other provinces being expressed relative to this province's deprivation index.

Table 3: Normalising deprivation index values and weighting the population

Province	Deprivation Index value	Normalised deprivation index score	Population	Population weighted by deprivation	% share of weighted population
A	0.86061	3.04954	6,302,525	19,219,802	22.0
B	-0.49656	1.69237	2,633,504	4,456,863	5.1
C	-1.18893	1	7,348,423	7,348,423	8.4
D	0.28813	2.47706	8,417,021	20,849,466	23.9
E	0.08357	2.2725	2,800,711	6,364,616	7.3
F	-0.604	1.58493	840,321	1,331,850	1.5
G	1.02013	3.20906	4,929,368	15,818,638	18.1
H	0.08551	2.27444	3,354,825	7,630,348	8.7
I	-1.12222	1.06671	3,956,875	4,220,838	4.8

The next step is to multiply the population in each geographic area by the normalised index value to estimate the weighted population. Thereafter, each geographic area's share of the weighted population is calculated (see Table 3 above). This deprivation weighted population estimate could serve as one component in an overall resource allocation formula. Other possible formula components could include each geographic area's unweighted population share, the relative areas' shares of population weighted for age and sex composition and the shares of standardised mortality. A range of other possible formula components have been suggested in the literature; the selection of formula components within each country should be based on what is considered to be appropriate within that context as well as data availability considerations.

If one is only using population size and deprivation in the resource allocation formula, each geographic area's percentage share of the population, weighted for deprivation, can be used as the basis for calculating target equitable shares of budgetary resources. As can be seen in Table 4, if there is a total budget of $259.9 million available for distribution to different provinces, the target equitable share per province is the total budget multiplied by the percentage share of the weighted population in that province. For example the equity target share of the budget for Province A is $259.9 million x 22%.

As can be seen from Table 4, there are often substantial inequities in the distribution of public sector health care resources between geographic areas. For example, Provinces C and I currently have a budget share that is substantially above their equity target share (despite being the least deprived provinces – see Table 3), while Provinces A and G have budget shares that are significantly below their target shares (yet are the most deprived provinces – see Table 3). Given the extent of the relative over- and under-funding of different provinces, the equity target budgets cannot be achieved overnight. A process of gradual redistribution of resources from relatively over-resourced areas to relatively under-resourced areas is required.

Table 4: Calculating equity target shares of budgets

Province	% share of weighted population	Equity target share of budget $ thousands ($ per capita)		Actual budget $ thousands
A	22.0	57,250	(9.10)	39,990
B	5.1	13,280	(5.04)	17,730
C	8.4	21,890	(2.98)	65,640
D	23.9	62,110	(7.38)	51,770
E	7.3	18,960	(6.77)	12,240
F	1.5	3,970	(4.72)	4,060
G	18.1	47,120	(9.56)	23,050
H	8.7	22,730	(6.78)	16,230
I	4.8	12,570	(3.18)	29,160
TOTAL	100	259,870		259,870

It is important to recognise that the underlying assumption of the method illustrated in Table 4 is that resources should be allocated in the same ratio as the relative levels of deprivation. For example, the level of deprivation in Province A is regarded as being 3.05 times greater than in Province C (based on the normalised deprivation index in Table 3). This translates, in the example below, into a target equity allocation where for every $1 spent per person in Province C, $3.05 should be spent per person in Province A and $3.21 in Province G (see per capita equity targets in Table 4). There is no empirical basis for this assumption, and there should be debate and policy engagement around how relative deprivation should be taken into account in resource allocation. For example, policy makers may believe that in order to reduce the relative differences in deprivation between provinces, about $6 should be spent in Province A and G for every $1 spent in Province C. It is important not to regard a resource allocation formula simply as a decision making ‘machine’, but as a tool that can assist policy makers and which can be adjusted to take into account public and/or policy maker preferences for the extent and pace of addressing relative deprivation.

Final comment

This booklet provides some ideas on how deprivation and resource allocation issues can be explored within the context of small area analyses. It provides step-by-step suggestions on how such analyses can be undertaken, based on the methods used in a study in South Africa, and it introduces some of the technical aspects of these analyses. It should not be viewed as a ‘recipe book’ that should be strictly followed, but as a resource that may provide insights into how researchers can undertake similar studies, with appropriate adaptations to meet each country’s needs and preferences. We hope that it is of value to those interested in researching and engaging in advocacy around deprivation and its relevance to government resource allocation decision-making.

appendix a

BRIEF BIBLIOGRPAHY FOR TECHNICAL ASPECTS

Additivity and weighting

Folwell K (1995). Single measures of deprivation.
Journal of Epidemiology and Community Health; 49 (Supp 2): S51 – S56.

Gordon G (1995). Census based deprivation indices: their weighting and validation.
Journal of Epidemiology and Community Health; 49 (Supp 2): S39 – S44.

Transforming variables

Gilthorpe MS (1995). The importance of normalization in the construction of deprivation indices.
Journal of Epidemiology and Community Health; 49 (Supp 2): S45 – S50.

Undertaking a PCA

Tabachnick BG & Fidell LS (1996). *Using Multivariate Statistics* 3rd Ed. Chapter 13 – Principal Components and Factor Analysis.

appendix b

STATISTICAL COMMANDS TO CONDUCT SUCH ANALYSES

1. Combining datasets

If you have two or more datasets that have the variables you wish to use for your deprivation study, it is important before you combine them to ensure:

1. That the same small area is used in both datasets, for example a district, and the districts are coded in the same way with the same number referring to the same district in each dataset; and
2. That both datasets are sorted by the small area code in the same way

<u>IN SPSS</u>

With one of the datasets you wish to combine showing in your data editor and ensuring the above, go to:

DATA → MERGE FILES → ADD VARIABLES

A dialogue box will appear to select the other file you wish to combine with. Open this external file.

Another dialogue box will then appear entitled Add Variables from [filename…]

Remove any variables from the window “New Working File” that you do not wish to appear in your final combined dataset. They will then appear in the “Excluded Variables” window.

Tick the “Match cases on key variables in sorted files” and select your small area code variable as your “Key Variable” by moving it from your “Excluded variables” window to “Key Variable”. You should also tick that both files have keyed variables.

Click PASTE.

Go to your syntax file and run the Merge Files set of commands.

A new combined dataset should then appear in your data processor window. Check for any missing data that you did not expect to be missing, to ensure that the merge was completed successfully.

IN STATA

In the command window run:

joinby [small area variable label] using [filename of the external file you wish to combine data from]

2. Undertaking the PCA

IN SPSS

In the data processor go to

ANALYZE → DATA REDUCTION → FACTOR ...

In the dialogue box that appears

Transfer the variables that you wish to use in the principal components analysis to the Variables box.

There are various options that need addressing, each of which has a button at the base of the dialogue box

Extraction

- Ensure the method showing is principal components (this is the default)
- Ensure show unrotated solution is checked (this is the default)
- Click scree plot (the scree plot shows you how many components should be considered as important. The point at which the line gradient substantially changes is this number of factors (this is usually between 2 and 4).
- Ensure that extract eigen values over 1 is checked (this is the default)

Rotation

- Check the varimax rotation box

Scores

- Check the save as variables box
- Check the display factor score coefficient matrix box.

Running the programme

Click paste and then run the DATA REDUCTION command section from your syntax file by highlighting it and then clicking on the arrow in the toolbar.

The Output

You will get output in your output file that should include the following:

1. correlation matrix
2. the unrotated solution
3. a scree plot
4. a table of the eigen values and the proportion of the variance explained by each factor and cumulatively
5. the rotated solution which gives you your components and each variable's factor loading
6. the matrix of factor score coefficients from the rotated solution

IN STATA

Two separate commands must be processed in Stata to complete a PCA (three if using another form of factor analysis).

factor [specify list of variables that the PCA is going to be performed on], pc factors(#) mineigen(#)

This will produce # number of factors (or components) as specified in the factors subcommand where the minimum eigen value is (#) as specified by the mineigen subcommand in the results window.

**** NB – all your commands and results should be saved in a log file ****

score [names of the variables you want factor scores to be called for each component]

Note that the default method of producing factor scores is the regression method and this is preferred.

This will produce a vector of factor scores for each component as retained in the factor command above. The scores will then be retained in the data under the variable name you have specified in the score command above

It is these factor scores that are then multiplied by their respective standardized variable to produce the composite index of deprivation.

EQUINET
methods toolkit

Equity in health implies addressing differences in health status that are unnecessary, avoidable and unfair. In southern Africa, these typically relate to disparities across racial groups, rural/urban status, socio-economic status, gender, age and geographical region. EQUINET is primarily concerned with equity motivated interventions that seek to allocate resources preferentially to those with the worst health status (vertical equity).

EQUINET seeks to understand and influence the redistribution of social and economic resources for equity oriented interventions, EQUINET also seeks to understand and inform the power and ability people (and social groups) have to make choices over health inputs and their capacity to use these choices towards health.

EQUINET implements work in a number of areas identified as central to health equity in the region:

- Public health impacts of macroeconomic and trade policies
- Poverty, deprivation and health equity and household resources for health
- Health rights as a driving force for health equity
- Health financing and integration of deprivation into health resource allocation
- Public-private mix and subsidies in health systems
- Distribution and migration of health personnel
- Equity oriented health systems responses to HIV/AIDS and treatment access
- Governance and participation in health systems
- Monitoring health equity and supporting evidence led policy

EQUINET is governed by a steering committee involving institutions and individuals co-ordinating theme, country or process work in EQUINET: Rene Loewenson, Godfrey Musuka TARSC Zimbabwe; Firoze Manji Fahamu UK/SA; Mwajumah Masaiganah Peoples Health Movement, Tanzania; Itai Rusike CWGH, Zimbabwe; Godfrey Woelk University of Zimbabwe, TJ Ngulube CHESSORE Zambia; Lucy Gilson, Centre for Health Policy South Africa; Di McIntyre University of Cape Town HEU South Africa; Gertrudes Machatini, Mozambique; Gabriel Mwaluko, Tanzania; Adamson Muula, MHEN Malawi; Patrick Bond Municipal Services Project; A Ntuli, Health Systems Trust, South Africa; Leslie London UCT School of Family and Public Health South Africa; Yash Tandon/ Riaz Tayob SEATINI, Zimbabwe.

For further information on EQUINET please contact the secretariat:
Training and Research Support Centre (TARSC)
47 Van Praagh Ave, Milton Park, Harare, Zimbabwe
Tel + 263 4 705108/708835 Fax + 737220
Email: admin@equinetafrica.org
Website: www.equinetafrica.org

DTP: Blue Apple Design, Bulawayo, Zimbabwe
Printing: Hunyani Printopak, Bulawayo, Zimbabwe