Breaking down barriers: Deprivation mapping at a local scale

Field, K

Abstract

In the UK, published national indices of deprivation are routinely used for resource allocation at a local level yet are constructed from national datasets. Data are usually derived from aggregate population based census sources using comparable geographical frameworks. This can ignore locality specific socio-economic and environmental factors associated with deprivation such as the location of crime hot spots, road traffic accidents or air pollution levels. The methodology developed here allows the combination of data collected using different spatial frameworks, from a range of sources and at both individual point level and aggregate area level to create a flexible approach to mapping deprivation at the local level. Deprivation grids are calculated using a combined statistical and GIS approach to portray deprivation at the local level in the county of Northamptonshire, UK. This approach has been adopted by Northamptonshire County Council to augment national indices to improve local decision making processes.

Introduction

The origins of this study stem from a realisation by the Overview Committee Working Group of Northamptonshire County Council (NCC) during March 2003 that national indicators of deprivation did not necessarily reflect reality at a local level. Given national policy objectives set out by the government to narrow the gap between deprived neighbourhoods, this gave cause for concern that current methods of identifying deprived neighbourhoods may not be sufficiently powerful enough. Several indexes of deprivation have been created at the national level for use, for example, by health authorities and local authorities. All are based upon the calculation of a 'national' average to which any geographical area (district, health authority, for example) can be compared. Using these national indices of deprivation, such as the 2004 Index of Deprivation (ID2004), South Northamptonshire District usually performs well and Corby Borough poorly (Figure 1). It is clear that this is as much a function of the geographical scale of the spatial units of analysis and the use of a national comparator as it is of the realities of advantage and disadvantage as they appear in both these districts at the local level. Sharp socio-demographic variations between the inner wards of the town of Corby and its neighbouring rural villages are 'lost' in an average figure for the borough as a whole. A similar averaging for a prosperous district such as South Northamptonshire inevitably hides pockets of disadvantage revealed when more locally based data sets are used. Although such measures can be extremely useful in allowing a county-based authority to see how it performs in a national context, they may not be sufficiently sensitive at picking up local variations. Even at super output area level, pockets of deprivation can be masked and may be balanced by the characteristics of the population elsewhere. In a policy sense this has real implications given the use of deprivation scores to assist in targeting priority areas for action and intervention and allocating financial resources. The research reported here develops a methodology for the determination of deprivation at a local level using the county of Northamptonshire, UK as a case study.

Underpinning the development of the methodology is a need to gather data at such a scale as to recognise the detail of local variations relative to a county standard. All national indices of deprivation developed so far have a major limitation in that they are largely tied to the availability of statistical data at a particular geographical scale, usually the administrative district or ward. Whilst this gives a degree of geographical consistency to the methodology, it also adds a high degree of inflexibility. By definition, geographical units of administrative convenience often create social boundaries on the map where none exist on the ground and often hide socio-demographical variations within an area by the calculation of an 'average' for the area as a whole.

Whilst it is evident that such arguments can be levelled against any spatial framework for the collection of data, the use of a fixed geographical unit eliminates the tapping of extremely useful data sets which have been gathered to other spatial frames. There are very useful data sets available which have not been used in earlier indexes of deprivation whether developed in the county or elsewhere. These include data gathered by police beat

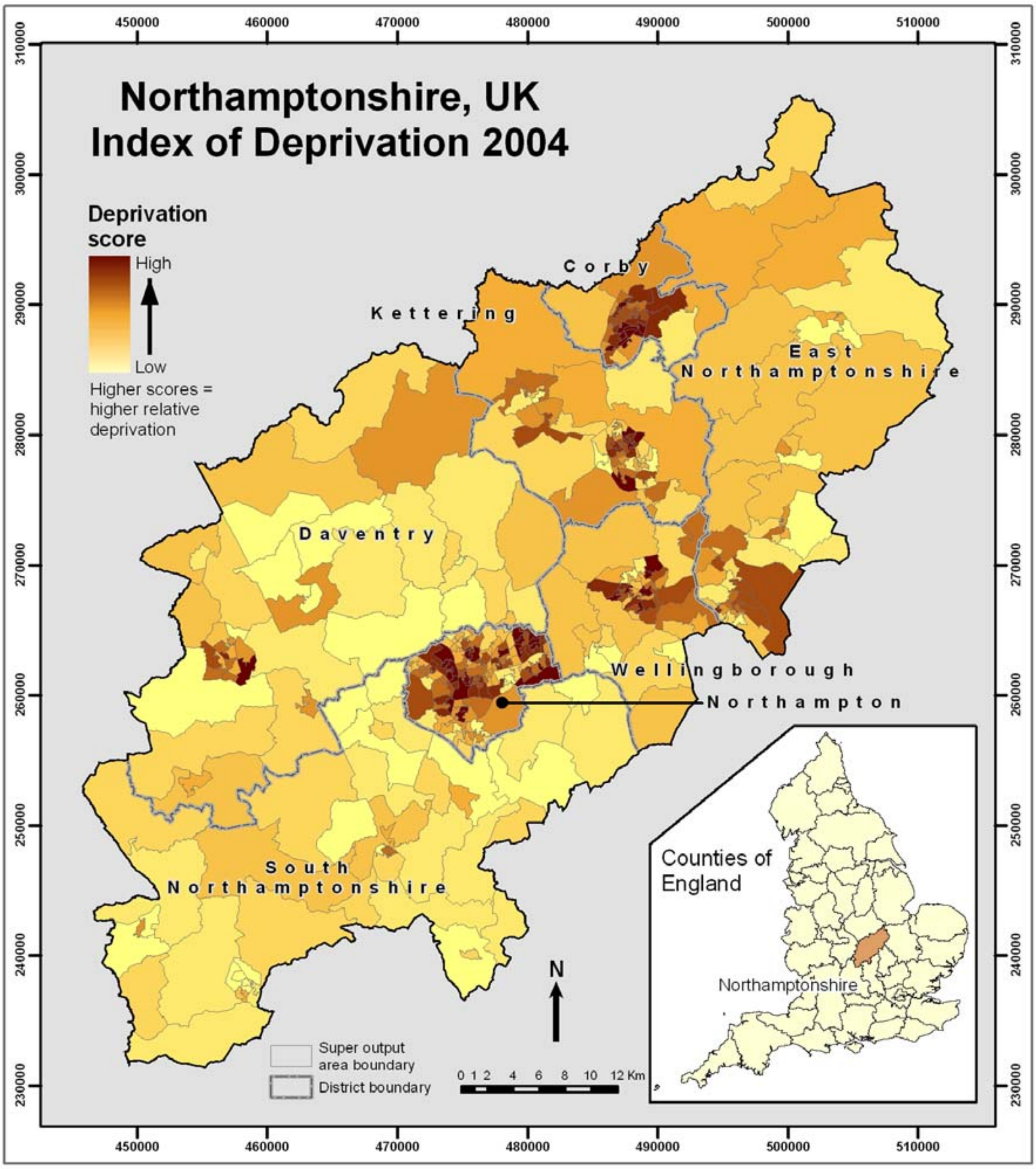

Figure 1. Index of Deprivation, 2004 for Northamptonshire, UK

areas, postcodes and point based sampling frameworks that may reflect more accurately local conditions.

There is a further aspect to the need for flexibility in the development of the methodology, namely the ability to be able to modify the constituent deprivation indicators as required and make them relevant to particular objectives. The factors that may be relevant in identifying areas which should be targeted for educational improvement may be somewhat different to those that identify areas of relatively high criminal activity. It is a limitation of many indexes that they become 'fixed in aspic' until replaced or updated by the availability of more recent data sets. In many instances the key data set is the decennial census, which provides a snap shot of the population structure and distribution for an area. As such, it degrades in value over time but whilst the importance of the census to any study of deprivation is recognised, it is essential that the methodology can be rolled forward to incorporate data sets as they are published, locally or nationally, whether to augment or replace data sets in the original model. In this sense, the local model of deprivation developed here should have the flexibility to evolve through time or to be adapted and tuned to the needs of particular groups and organisations.

Identifying indicators of deprivation

In the UK, researchers and policymakers have made numerous attempts to measure deprivation that explore alternative units of geographical analysis and statistical treatments of the data as well as proposing indices based on a wide range of indicators and for different purposes (Jarman, 1983; Townsend, 1988; Carstairs and Morris 1989a, 1989b; Carr-Hill and Sheldon, 1991; Field, 2000; Noble *et al.*, 2000; ODPM, 2004). Since the 1960s, the identification of priority areas for targeting particular programmes of improvement has been a feature of government policy. The development of a national index of deprivation has been refined over the past two decades with various indices being published, most notably 1981’s Index of Deprivation, 1991’s Index of Local Conditions, 1998’s Index of Local Deprivation, 1999’s Index of Multiple Deprivation and, most recently, the 2004 Indices of Deprivation (ID2004). These have employed a range of alternative approaches to the construction of an index in geographical and statistical terms. The 1981 index, for instance, used the sum of the z-scores of the component items whereas the 1991 version employed a signed chi-squared statistic determination. National indices have also increasingly moved towards the use of direct measures of deprivation, the use of ‘domains’ to classify aspects of deprivation, the incorporation of datasets other than the decennial census and the calculation of scores at small area level. These are all positive steps and with each iteration of the method, it becomes increasingly sophisticated and more robust. However, as recently as the 1999 review of the Index of Multiple Deprivation was recognition that the use of small geographical areas remained problematic given boundary alterations, lack of data available at that scale and the reliance on estimation of denominators (Noble *et al.*, 1999).

There are a number of important aspects to address in creating any index namely to determine:

- what the conceptual basis of measurement should be;
- what are the indicators of deprivation;
- how weighting of component indicators might be employed;
- how are the spatial scales of input data related;
- in what way can areas of different sizes be accommodated; and
- what is the smallest level of spatial aggregation for the index.

Deprivation cannot be measured on its own. The term deprivation is most commonly applied to people who are lacking in resources and other conditions of life based on financial or social circumstance. It does not exist in its own right and cannot be measured so there is a reliance on indirect measurement using a combination of component indicators. Since the methodology was to enable development of a local measure of deprivation for Northamptonshire, a set of indicators were constructed and agreed in consultation with officers and members of the various Scrutiny Committees of NCC via a questionnaire survey. This sought to establish issues that are of local concern and the extent they impact upon the workload of the Council. For instance, the Health and Social Care Scrutiny Committee reported the following factors as being of local concern which might reflect patterns of health and social related deprivation:

- Age specific alcoholism and alcohol dependency
- Young People and Substance Abuse
- Older People in Rural Areas
- Circulatory disease mortality for those age <75
- Low life expectancy

Seeking indicator choice at the outset ensured that decisions were not constrained by data type, data source, geographical or temporal framework since the overarching objective of this research is to develop a methodology that can handle disparate data. Local indicators were established that mirrored the 'domains' established in the IMD2000 (and subsequently, the ID2004 although this had yet to be published at the time the research reported here was in progress) to provide a mechanism for locally adjusting national measures of these domains at a later stage. Table 1 sets out the indicators chosen through consultation with NCC. Data was obtained from published sources where available or provided by the offices of NCC. A wide variety of data was identified to reflect local conditions in Northamptonshire.

The potential interaction of component indicators has led some to weight certain indicators, when combined, more importantly than others (e.g. Jarman 1983). This approach seeks to identify the relative impact of certain factors prior to combination in a composite index. There is, of course, an argument that equal combination of component indicators is a form of weighting and that prior weighting may mask nuances in the data and for certain areas. The questionnaire survey sought to determine whether there were any particular factors that were particularly indicative of deprivation in Northamptonshire to inform weighting. Response was varied and given that different stakeholders put forward competing views, according to their particular interests, it was decided not to weight component indicators when combined.

Whilst the methodology developed here allows the analysis of data in different measured units and for different geographical areas, it is essential that for any individual indicator that the unit and area are consistent. In other words a data set was rejected where, for example, it was available for wards for six of the county's administrative districts but only as a district figure for the seventh.

Domain	Indicator	Temporal framework	Spatial framework
Health	% Limiting Long Term Illness	2001	2001 ward
	Average life expectancy (male and female)	1995-99	1991 frozen ward
	Alcohol and drug dependancy ?	2003	Postcode district
	% deaths from circulatory disease age <75	1995-99	1991 frozen ward
	% deaths from cancer age <75	1995-99	1991 frozen wards
Income	% Income support claimants	2000	1991 frozen ward
	% Family tax credit claimants	2000	1991 frozen ward
Work deprivation	% age 16-74 in Elementary occupations	2001	2001 ward
	% Claimant unemployed <25	2003 (1991 frozen ward)	1991 frozen ward
	% Claimant unemployed > 12 months	2003 (1991 frozen ward)	1991 frozen ward
Social	% teenage pregnancies	1998-2000	2001 ward
	recorded hate incidences	2002-3	Police beat
	recorded drug offences	2002-3	Police beat
Housing	% Dwelling stock Bands A-B	2001 (1991 frozen ward)	1991 frozen ward
	%households with occupancy -1 or less	2001	2001 ward
Crime	Malicious fires	2002-3	Unit Postcode
	Number of all recorded crimes	2002-3	Police beat
Education	% 18-19yr olds not proceeding to University	1998	1991 frozen ward
	% age 16-74 no qualifications or level 1 NVQ	2001	2001 ward
Access to services	% pensioner households with no car	2001	2001 ward
	% households > 2km from GP surgery	1998	1991 frozen ward
Physical environment	Carbon monoxide concentration	1998	OS grid coordinate
	Nitrogen dioxide concentration	1998	OS grid coordinate
	Child pedestrian accidents	2000-2003	Unit Postcode
Ethnicity	% Bangladeshi male unemployment	2001	2001 ward
	% people in Bangladeshi households with occupancy rating -1 or less	2001	2001 ward
	% afro caribbean 16-24 unemployed	2001	2001 ward

Table 1. Local indicators of deprivation for Northamptonshire

Similarly a data set was rejected if it was available at ward level throughout the county but where the form of measurement differed between one district and another.

Where possible the data sets used were those available at the smallest geographical scale consistent with the statistical usefulness of the data at that scale. Throughout the study no geographical unit larger than the census ward has been used, of which there are currently 150 within the county, so for data derived from the 2001 census each indicator is a function of a minimum of 150 real values for the county. In most cases far more real values have been used, for police beats or discrete point based events (e.g. malicious fires) for example, but only where the data are meaningful at that level.

Developing a local index of deprivation: spatial and statistical treatment

As the background and developmental discussions that precede this section have highlighted there are three main issues that remain to be tackled. These are that:

- locally relevant data is reported using different spatial frameworks;
- different datasets are collected for different time periods; and
- values and distributions of data differ between datasets.

These issues lie at the heart of attempts to create a composite measure of multiple deprivation from component indicators. They have presented difficulties in preparing deprivation indices more widely and most have overcome the problems by using a single spatial framework (e.g. ward) or using data from a fixed source (e.g. a single census). Whilst there are sound reasons for taking this approach there is no doubt that it also restricts the range and type of data that might be included. It has been the purpose of this research to determine a methodology that is flexible enough to incorporate data that exist in a range of different spatial forms, that reflect a wider range of indicators of deprivation and that may have been collected at different points in time by different organisations and local stakeholders. It is therefore attempting to loosen the constraints of rigid, and sometimes arbitrary, geographical structures and to provide an opportunity to use data from a range of sources.

Nevertheless, it should be noted at the outset that this flexibility gives the methodology significant power to incorporate a much wider set of indicators than other indices but that the same standards of rigour should be applied to indicator selection. Clearly, data that is collected at points in time that vary considerably will be more prone to spurious interpretation. As with any exercise of this type, the more comparable and compatible the datasets in terms of temporal or spatial homogeneity then the more meaningful the outcome when they are combined.

In order to address the threefold difficulties highlighted, the methodology applies a range of Geographical Information Systems (GIS) techniques and statistical treatments to each dataset. Thus each dataset becomes spatially and statistically comparable with each other and they can then be aggregated to give a composite measure of deprivation. This process is illustrated in Figure 2 and described below.

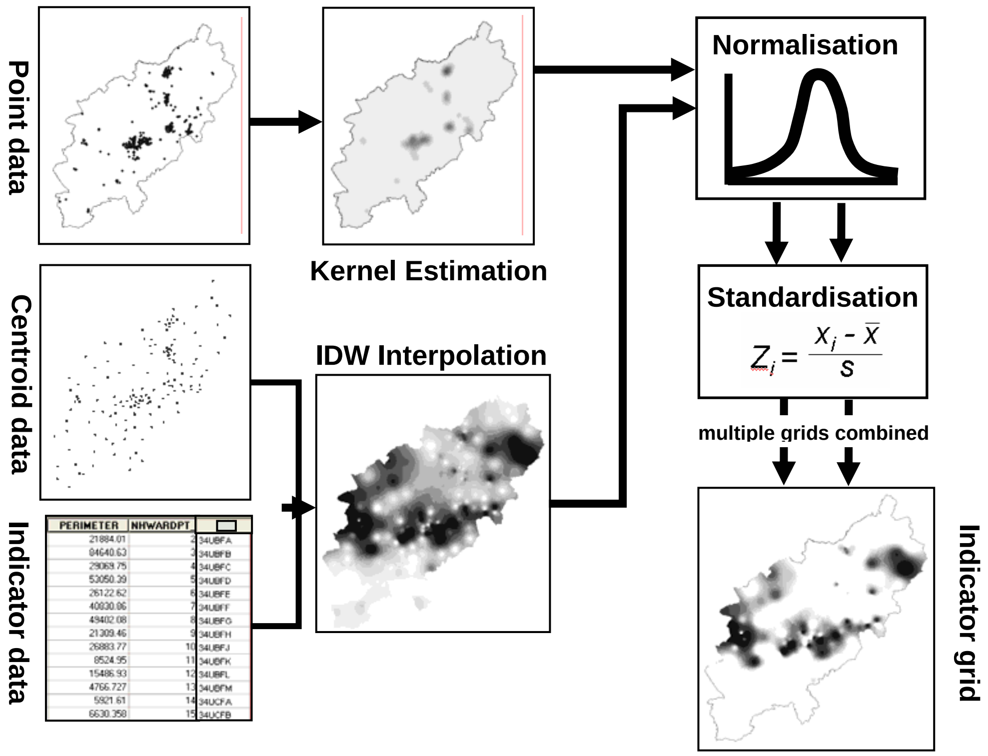

Figure 2. Flow of data manipulation in the methodology for determining deprivation

A wide range of geographical (spatial) datasets exist which are used to collect, analyse and present data. These may be areal in form (e.g. census ward boundaries, police beats) or point based (e.g. unit postcode centroids, Ordnance Survey grid coordinates). These are incompatible in their raw form so a preliminary decision must be made to determine how data will be transformed in order to provide a consistent spatial framework. Grids were chosen since they provide a sound mechanism for transforming data from any spatial framework to one that is not constrained by internal political, administrative or geographical boundaries. Grids do not change over time, can be fixed at a specific resolution and do not introduce any additional bias.

Of the indicators determined by the various consultation phases of this project (Table 1), the following spatial frameworks are used to represent individual indicators prior to them being transformed to grids:

- 1991 Census ward boundaries (polygon);
- 2001 Census ward boundaries (polygon);
- police beat boundaries (polygon);
- unit postcode centroids (point);
- Ordnance Survey grid coordinate specific data (point);and
- Ordnance Survey grid coordinate sampled data (point).

Three polygon datasets are used where data are reported as a single aggregated figure across a related geographical area. Whilst the most up-to-date indicator data have been used wherever possible it is has been necessary to make use of 1991 census ward boundaries for some datasets. These include those which have been collected and reported prior to the 2001 census boundaries (e.g. % income tax claimants, 2000) and those for which 1991 frozen boundaries have been used despite data being collected post-2001 (e.g. % claimant unemployed who are age <25, 2003). The 2001 census boundaries have been used for all data derived from the 2001 census (e.g. % pensioner households with no car). Data relating to incidence of crime is reported using police beat boundaries. These are largely co-terminus with aggregated census wards but must be used to accurately prepare these datasets (e.g. recorded hate incidences as a % of all crime, 2002-3).

The remainder of the datasets are point based and usually record a single incidence. The points themselves determine the pattern across geographical space and these datasets are rarely presented in aggregate form in geographical areas. This is a key feature of the methodology that deviates from others; namely, that point data measuring an indicator at a point location can be used in conjunction with the more commonly used area based datasets. One such dataset identified for this study is child pedestrian accidents where unit postcodes are used to provide a geographical reference to incidence.

Some datasets (e.g. malicious fires) are location specific and each incidence is reported along with a 6 figure Ordnance Survey grid reference, accurate to 100m. The remainder are sampled at points across geographical space (e.g.

Carbon Monoxide atmospheric concentration sampled at 1kmx1km and reported as a 6 figure OS grid coordinates).

As noted, the spatial frameworks of these datasets differ but they must be transformed to a comparable grid format to be combined. Areal interpolation techniques have been used in studies to reconcile data from incompatible spatial frameworks (Flowerdew and Green 1991, 1992) and is the method used here but interpolation relies on point based input data. The location specific and sampled point datasets are easily dealt with because point based location is the inherent in the OS grid reference reporting mechanism. Unit postcode areas (representing c.30 postal addresses) are represented by their polygon centroid and converted to their OS grid reference equivalents using OS code-point data.

The polygon-based datasets must also be converted into their point equivalents. GIS tools are used to extract the polygon centroid. Thus the 150 wards in the 2001 census ward boundary dataset are represented by 150 points in a new map layer. The position of the point in geographical space is a question to be considered since it will have an impact on the subsequent interpolation process and, ultimately, the final output. There are two choices for identifying the location of a centroid – either using the mean centre of the polygon in which it is to be placed (Figure 3a) or using a weighted mean centre based on underlying geographies and additional data. Figure 3b illustrates a weighted mean centroid adjusted spatially, towards populated areas rather than unpopulated (i.e. it provides a more accurate representation of the true location of the spatial distribution from which the population-based indicators have been measured).

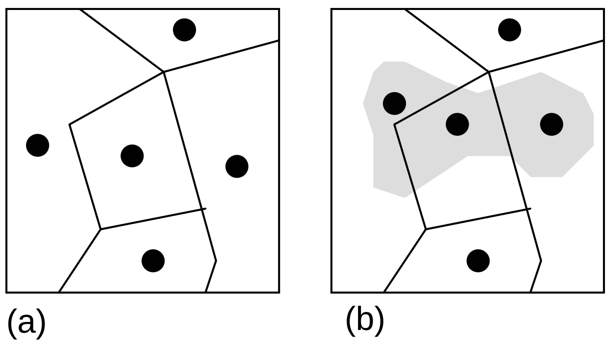

Figure 3. Alternative approaches to locating centroids in polggons

The choice made in developing this methodology was to use a weighted mean centre of each area in the polygon datasets according to the location of population concentrations. This was applied consistently for each of the three polygon datasets.

Measuring indicators using totals are avoided, particularly where data items represent an aggregate and are normally reported via an area based spatial dataset. The reason for this is that the size of the area will have a major impact on the numerical value. This is commonly referred to as the Modifiable Areal Unit Problem (Openshaw, 1984), a problem that affects any representation of spatial data in area based form. Whilst the constraints of areas have been partially tackled by converting the spatial datasets to points, nevertheless, each data value should be expressed in its percentage form where feasible. For example, simply using the number of pensioner households that do not possess a car as an indicator fails to acknowledge whether this figure is high or low relative to other areas in the study area because no account is made of the population base that the values emerges from. In this form, it does not give a basis for determining relative levels of deprivation. The only way to remedy this is to express the value as a % of the appropriate denominator (in this case all households) in each area and create a measure that can be compared across the map. Each of the datasets are thus manipulated to ensure the data values are expressed in appropriate terms.

For data collected at location specific and sampled points these processes are not required since the absolute value is measured uniquely at points. The data are not aggregated from finer resolutions or representative of populations bounded by an arbitrarily defined geographical area. Thus the absolute value can be used for areal interpolation of point based datasets (e.g. incidence of malicious fires).

Areal interpolation

A number of alternative interpolation techniques may be used to create (interpolate) a continuous surface. All methods are based on examining the similarity of nearby points to create the surface. Deterministic techniques use mathematical functions for interpolation such as Inverse Distance Weighting (based on the extent of similarity between nearby points) or Radial Basis Functions (based on the extent of smoothing). Both of these techniques are local interpolators, that is, they use a local neighbourhood search to define an area in which points will be used in the calculation. This is based on the principle that points that are closer together will have a greater influence than those farther apart.

Deterministic global interpolators also exist (e.g. global polynomial interpolator) that make use of the entire dataset in the interpolation calculations. The choice of which interpolator to use is driven by the purpose of interpolation. In the case of the datasets used here, a technique is required that allows a continuous surface to be created to represent the point-based data representations for each of the indicators. Global interpolators are not appropriate since variation in the indicator is much more likely to be locally determined and a local interpolator is more appropriate. A local interpolator may either force the continuous surface to pass through the measured points (an exact interpolator) or to vary from the actual value at each measured point (inexact interpolator).

Inexact interpolators are usually used to smooth data across a surface and lessen the impact of peaks and troughs. However, extreme values may be important in the identification of deprivation hot spots so an exact local interpolator is the most appropriate method to use. Of those that are commonly available in GIS software, the Inverse Distance Weighted (IDW) is robust and has fewer demands on the dataset than other methods and is the method adopted in this study.

For each of the indicators, a grid is interpolated. A variable distance search radius was used and interpolation was defined by the inclusion of the twelve nearest points. The final consideration is the resolution of the output grid cells. A low resolution grid would have large cells, say $1km^2$ but, of course, a considerable amount of generalization of the surface would take place as part of the interpolation, particularly in areas with many nearby points. A high resolution grid of, say, $1m^2$ would provide a very fine level of detail but is inappropriate due to it connotating a level of accuracy not inherent in the data used and also resulting in very large data files. A resolution of $100m^2$ was established for all grids calculated in this research to be consistent with the level of OS grid coordinate data used. Grid calculations in GIS output a rectangular grid of *n* rows by *n* columns. For the data in this study the grids, at a resolution of $100m^2$, were 748 rows by 670 columns resulting in grids with 501,160 separate values of interpolated data. Of course, a rectangular grid contains cells that fall outside the study area and are not, therefore, meaningful. Each indicator surface is clipped to the shape of the study area so that only grid cells within the study area are used in subsequent statistical manipulation.

The process of interpolation is slightly different for those datasets that are point specific (e.g. malicious fires). In these cases, the point dataset represents a single (e.g. malicious fires) or multiple occurrence (e.g. alcohol and drug dependency) of the indicator measured by a single mapped point. Thus the total number of points across the map represents both incidence and prevalence. A kernel estimation is performed on these datasets using a circular bandwidth of 2km applied to each cell and calculating a density value for each 100m cell in the output surface grid. This is based on the number of points within the search radius, the point value and the area of the search radius. The process weights the points near the centre of the search radius more than those at the periphery to incorporate a distance decay function. This provides a method of interpolating discrete values of point-based data comparable to the IDW interpolation applied to the polygon-based data.

Statistical Treatment

Regardless of the indicators chosen, each is statistically incompatible with another in their raw state. Bias will be inherent in any method which simply aggregates indicators in their raw state, due to differences in their scales of measurement, statistical range and relative importance to the outcome measures of interest. As has already been noted, weighting of indicators was not pursued in this study (although it should be noted that the method of construction would allow weighting to be incorporated at a later stage if necessary) so is not a further consideration here.

The statistical treatment of data for developing indices are largely based on one of two methods, the calculation of chi-square values or z-scores for each of the indicator datasets. The chi-square approach is more usually employed in large scale studies where geographical areas are at their most detailed (e.g. census output areas). It is also more appropriate when absolute values are used rather than percentage data and, thus, is not an appropriate approach for this study.

Alternatively, the z-scores approach first normalises data, to convert skewed distributions into normal distributions, prior to being standardised. This is usually achieved by calculating the logarithmic value of the data. The use of logarithmic values lessens the impact of skewed values, although, by definition, the range of values is reduced. However, the transformation still maintains the variance for any one variable and therefore those variables which illustrate greater variation still have the same impact when combined to create composite scores.

Once normalised, each variable is standardised to have a mean value equal to zero and a standard deviation equal to one (Eq 1)

$$Z_i = \frac{x_i - \bar{x}}{S}$$

[Eq 1]

This method provides results which are explicitly referenced to the overall mean for the study area as a whole. The resulting scores are interpreted thus: increasingly positive values are indicative of a higher relative deprivation. It should be noted, however, that the actual scores themselves are not in any way a reflection of any absolute measure of deprivation. They are a means to examine the relative levels of the indicator across the case study area. They are also only relative to the case study area and should not be compared like for like to any similar scores for other case study areas. Furthermore, it is worth noting that the indicators themselves have been chosen to reflect known dimensions of deprivation. That an area exhibits relatively low levels in comparison to another area does not in itself mean it doesn't suffer deprivation.

For the purposes of this study the base 10 logarithmic value was calculated for each indicator surface which provided a good transformation for all datasets. The results of these calculations create a new set of grid datasets representing each indicator which should then be standardized by calculating the z-score. This creates another new set of grids, one for each indicator. The flow of data manipulation is summarized in Figure 3
.

Validation of the methodology and results

Once all indicators have undergone the processing and analytical steps to create individual indicator grids, they are both spatially and statistically compatible. In order to construct a composite map of deprivation, the various indicator grids must be aggregated. This may be done in various ways.

Possibly the most commonly used is arithmetic aggregation: the indicators are either added or multiplied together to give a final index.

Additive combination of indicators takes no account of interactions between the indicators - for example, where the effect of the two acting together is greater than the effect of both, acting separately. Multiplicative aggregation implies interactions. Nevertheless, multiplicative aggregation does require detailed understanding of the relationship between indicators which can be complex and prone to uncertainty.

Given that knowledge of indicator interaction is unknown and difficult to quantify, the most appropriate method of aggregation is arithmetic. Since each indicator has been processed into a statistically normalised and standardised form they are numerically equivalent to one another and can be summed additively to produce a map of deprivation. The higher the score, the greater the relative deprivation exhibited. Large positive values suggest an area that is suffering multiple deprivation as the component indicators might all be at the higher end of the indicator score. Care should be taken in interpreting the composite deprivation maps since it is not possible to determine the interaction of individual indicators at any one point. Figure 4 illustrates a composite map of local deprivation for the County of Northamptonshire.

As stated, the purpose of designing a methodology to determine deprivation locally was to provide a mechanism to use data that is perceived to be more representative of local conditions. Furthermore, it has allowed a much more diverse set of indicators to be used since it has purposefully not been constrained to a single spatial framework. In order to assess the extent to which this new methodology can provide a predictive tool for assessing deprivation locally, some validation is required. This process also assesses the utility of the local methodology in relation to more established national frameworks for assessing measuring deprivation.

Correlation analyses

A valid dataset not used in the construction of the deprivation index, illustrating the location of all children in the County who have not achieved 5 or more grade C, or above, GCSE qualifications was used to test the methodology of measuring deprivation. The dataset provides information on the location of children with low levels of educational attainment, reported as a total number of children per unit postcode. The principle underpinning the assessment is that children with low levels of educational attainment are more likely to be resident in populations that display higher relative deprivation. Low educational attainment is simply one social manifestation of deprivation so the ability to be able to identify localities and hot-spots for priority action is important in targeting educational resources.

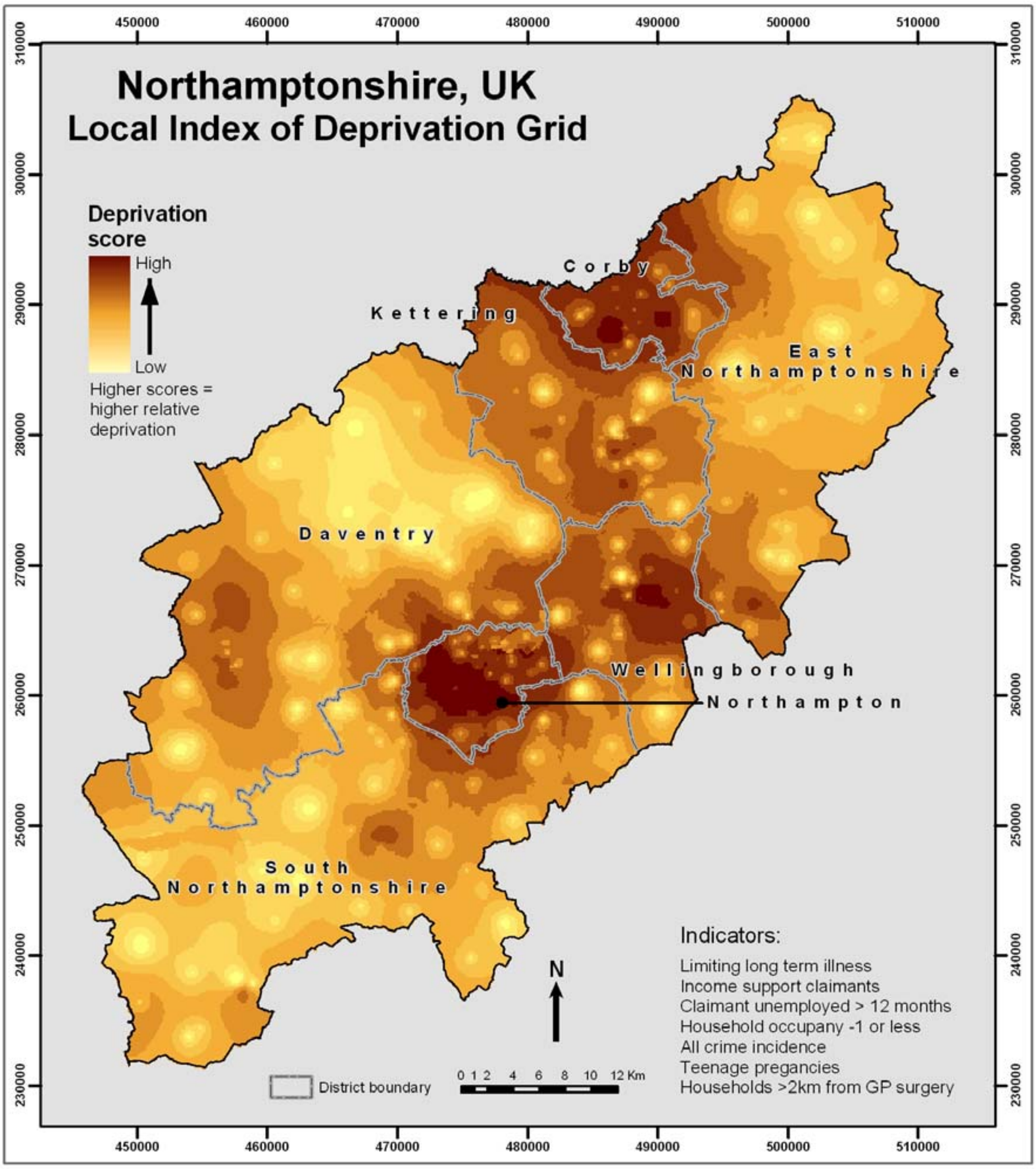

Figure 4. Composite grid of local deprivation for the County of Northamptonshire

Base map © Crown Copyright 2005. An Ordnance Survey/JISC supplied service

A kernel estimation was performed on the point-based low educational attainment dataset using a bandwidth of 5km in the kernel calculation to create a density surface of low educational attainment (Figure 5). The grid resolution is set to the same as the deprivation grids, namely 100m. The composite deprivation grid, comprising a complete range of deprivation indicators informed through consultation with NCC and indicative of factors that are thought to shape deprivation locally (Figure 4) was then correlated with the density surface for low educational attainment using Grid tools. This analysis yields a low positive correlation coefficient of r=0.58 (r^2=0.34).

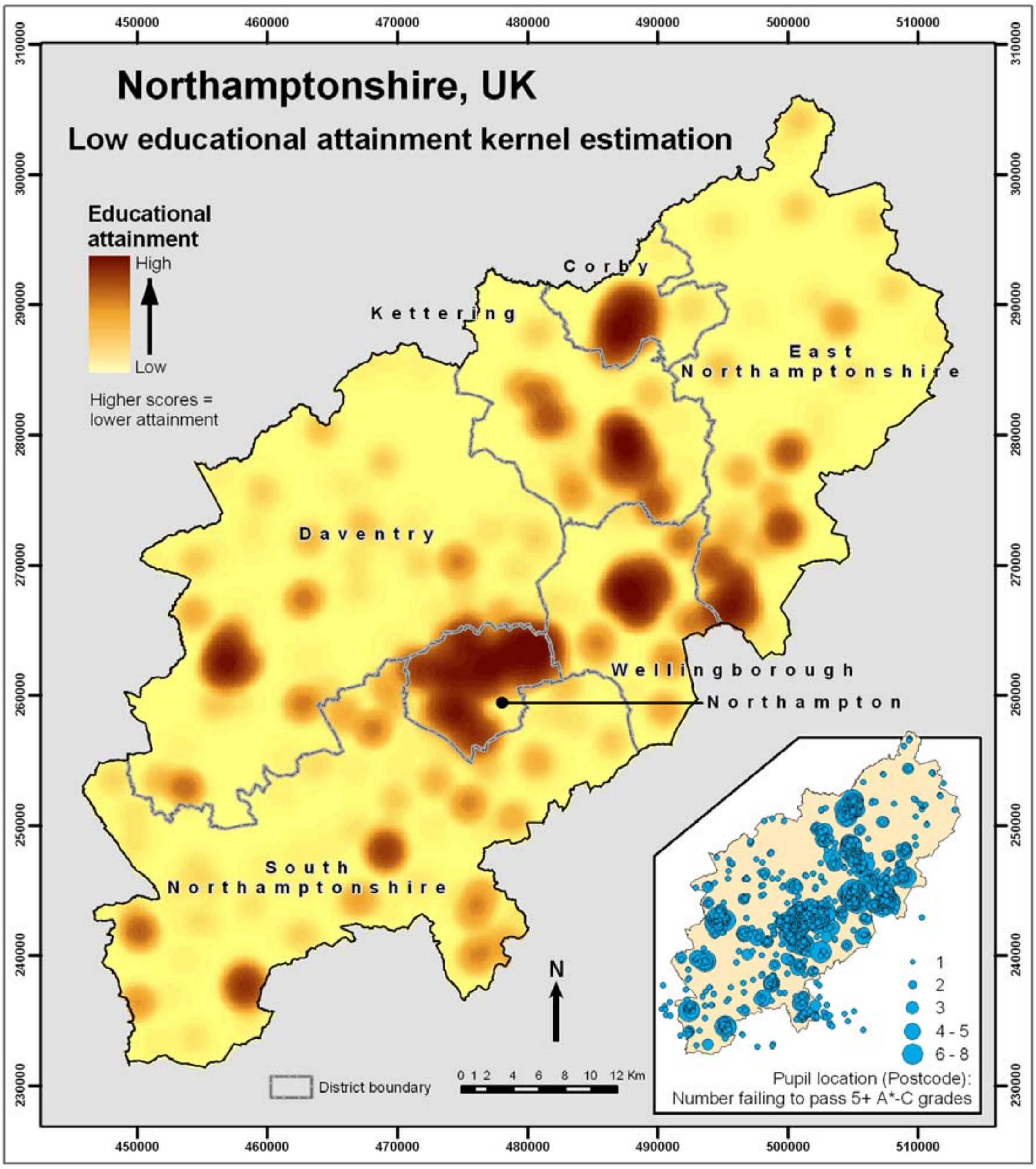

Figure 5. Kernel estimation of low levels of educational attainment for the County of Northamptonshire

A further correlation was performed between the low educational attainment density surface and a deprivation grid comprising fewer indicators that were identified by NCC as being particularly indicative of educationally related deprivation (Figure 6). This yielded a correlation coefficient of r=0.49 (r^2=0.24). Both of these analyses give low positive correlations and do not provide any significant evidence to suggest that the measures of deprivation developed to reflect local conditions actually perform adequately under statistical scrutiny.

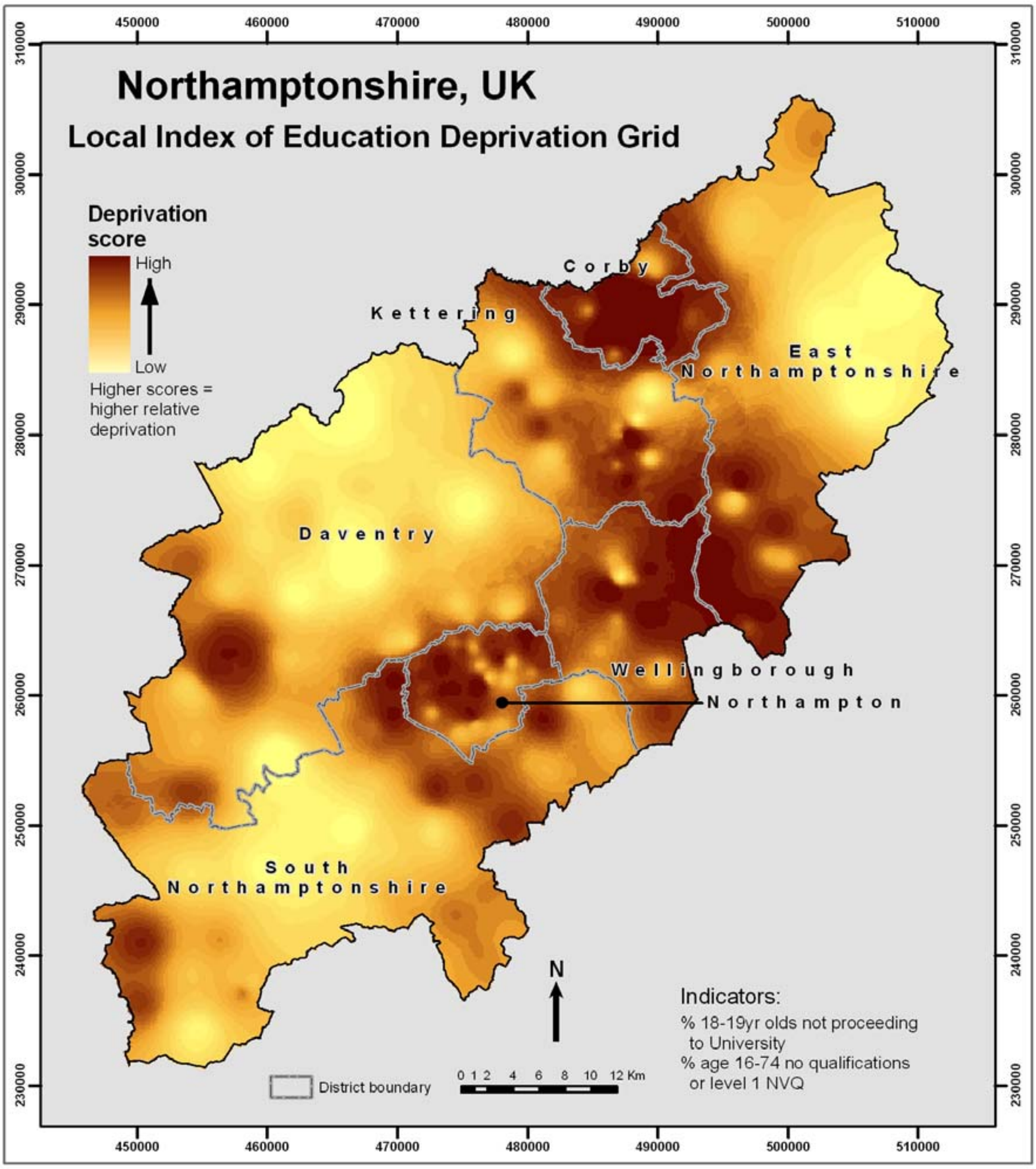

Figure 6. Composite grid of education deprivation for the County of Northamptonshire

As noted earlier, one of the goals of the work was to establish a methodology that was better able to reflect local conditions that were thought to be masked in national indices. The ID2004 is the most recently available national deprivation index, calculated for census super output areas (Figure 1). Population weighted centroids were used to transform the polygon-based data to a grid using the same parameters and IDW interpolation used to create the local deprivation grids. This ID2004 grid (Figure 7) was analysed in relation to the low educational attainment density surface giving a positive correlation coefficient of r=0.61 (r^2=0.38). Whilst the ID2004 can hardly be used as a

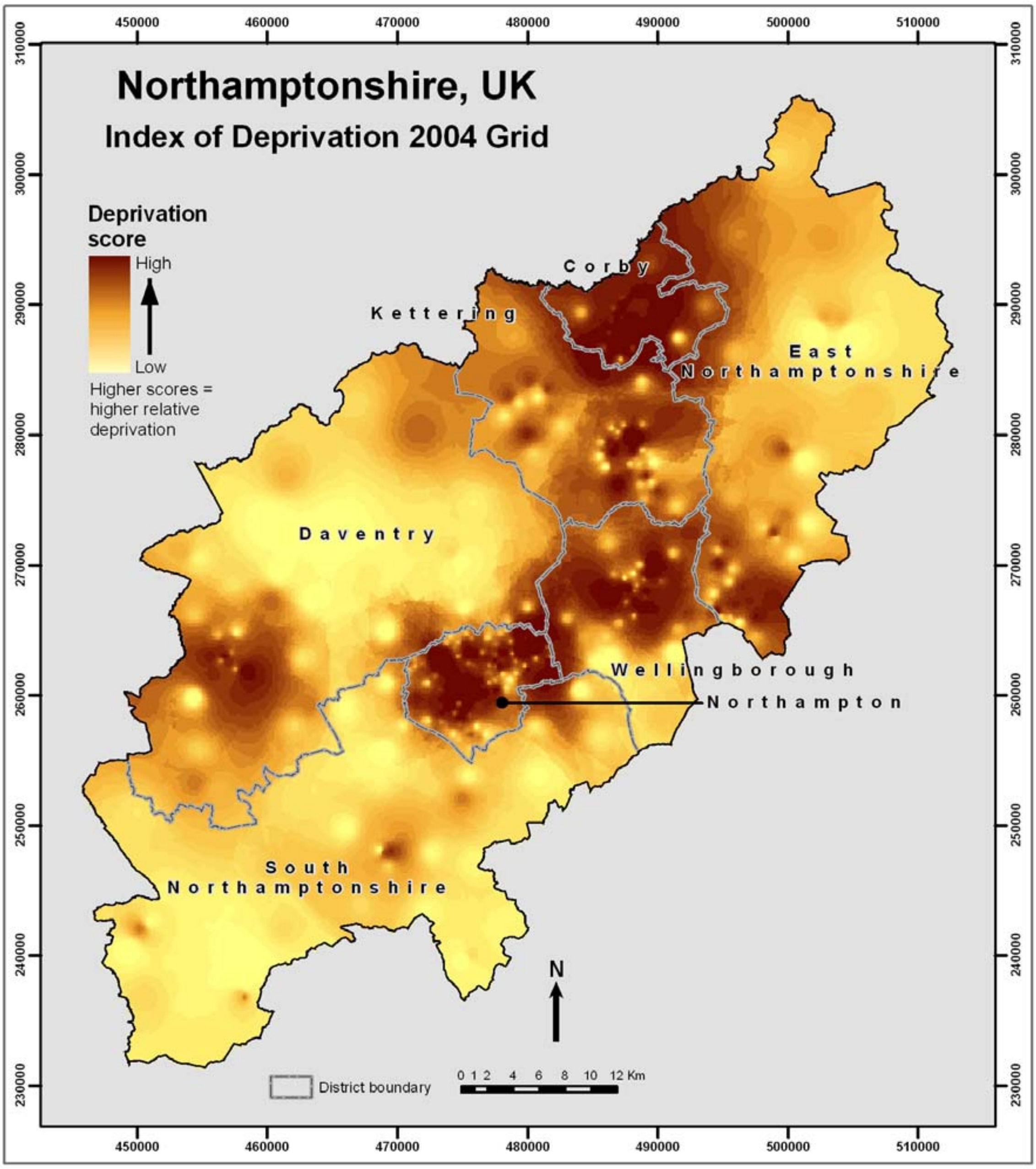

Figure 7. ID2004 Deprivation grid

reliable tool for predicting the spatial pattern of low educational attainment it, nevertheless, does illustrate that the national index is marginally more powerful than the locally derived indices. ID2004 is comprised of several domains specifically designed to measure certain aspects that explain patterns of deprivation, one of which is the education domain (Figure 8). This domain measure was similarly transformed to a grid using the same technique as above and gave a correlation coefficient of r=0.4 (r^2=0.16) when examined in relation to the low educational attainment grid. These correlation

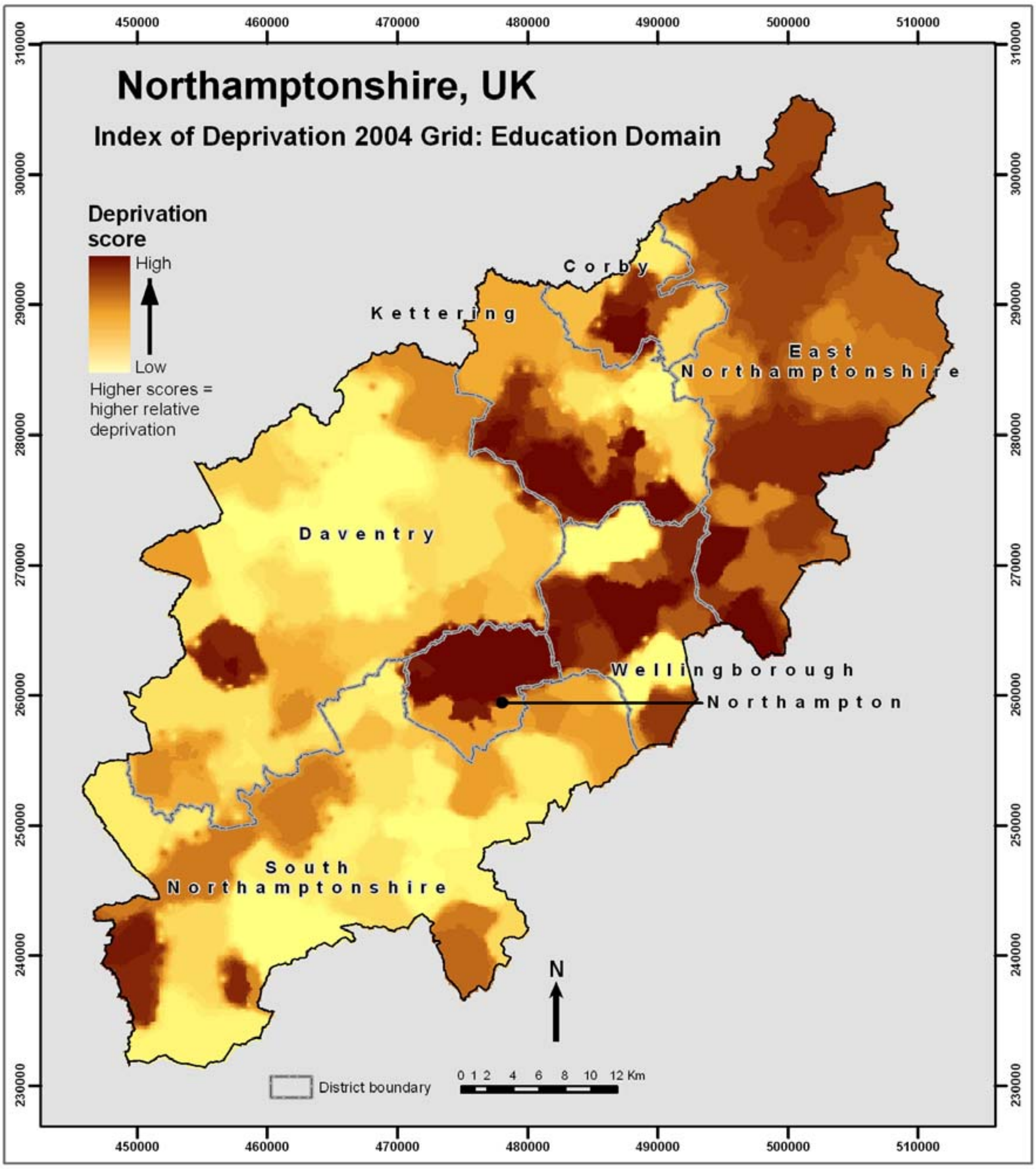

Figure 8. ID2004 Deprivation grid: Education domain

coefficients are no better than those for the local indicators. Unexpectedly, less variation in the lower educational attainment dataset can be explained by the education domain measure than the overall measure of deprivation provided by ID2004. It would have been expected that the education domain was better able to model the areas of low educational attainment but this is not the case for this case study. This suggests that there may be other aspects of deprivation that are exerting greater influence locally than those that are designed to measure education-related deprivation, meaning that the overall ID2004 is better able to predict areas of low educational attainment. Clearly this raises an important issue: that at the local scale, individual ID2004 domains may not be appropriate for identifying deprivation thematically due to the potential confounding influence of other factors that might be more powerful at the local level.

These results must be treated with some caution given the assumption that patterns of low educational attainment are assumed to be a reflection of wider patterns of deprivation in the population. Clearly, such an assumption takes little account of any potential confounding factors.

Conclusions

The methodology for determining local deprivation does not appear to provide any more predictive power than indices that are published and available nationally based on this case study area. As with any assessment of deprivation, the data that is included will shape the outcome and it may be the case that alternative, better or more powerful datasets exist. The purpose of the research has been to design a methodology which is local and flexible and which allows users to examine deprivation in their terms, whether focusing on a number of indicators in a particular category (e.g. education) or across a range of categories. In this sense, the methodology can be modified using alternative datasets to determine whether there are ways to improve on the national indices on a case by case basis.

The methodology is best seen as a tool, to be handled with care. This paper highlights a number of key factors which should be borne in mind when any datasets are combined spatially and statistically. It also emphasises the key decisions taken in determining appropriate indicators and deciding on the parameters for statistical and GIS based manipulation. Any results are largely a function of these decision making processes.

The methodology is currently being used by NCC in conjunction with national indices of deprivation to provide a range of tools capable of adding value to the decision-making process and location allocation of resources.

Acknowledgements

The author is grateful for the assistance of the various members of Northamptonshire County Council and the Scrutiny Committees in developing this work and to colleagues Dr Ken Sherwood and Jan Jones.

References

Field, K. S. (2000) Measuring need for primary health care: an index of relative disadvantage, Applied Geography, 20 (4) pp305-335.

Flowerdew, R. and Green, M. (1991) Data integration: statistical methods for transferring data between zonal systems. In Handling Geographic Information: Methodology and Potential Applications edited by I. Masser, and M. Blakemore (Harlow: Longman), pp38–54.

Flowerdew, R., and Green, M. (1992) Developments in areal interpolation methods and GIS. Annals of Regional Science, 26, pp67–78.

Jarman, B. (1983), Identification of underprivileged areas, British Medical Journal, 286, pp1705-1709.

Townsend, P., Phillimore, P. and Beattie, A. (1988) Health and deprivation: Inequality and the North. Croom Helm, London.
Carstairs, V. and Morris, R. (1989a) Deprivation and mortality: an alternative to social class? Community Medicine, 11, pp210-219.
Carstairs, V. and Morris, R. (1989b) Deprivation: explaining difference in mortality between Scotland and England and Wales. British Medical Journal, 299, pp886-889.
Carr-Hill, R. and Sheldon, T. (1991) Designing a deprivation payment for GPs: The UPA (8) wonderland. British Medical Journal, 302, pp393-396.
Noble, M., Penhale, B., Smith, G. and Wright, G. (2000) Index of Deprivation 1999 Review: Measuring multiple deprivation at the local level, Department of the Environment, Transport and the Regions (Online at: http://www.odpm.gov.uk) Accessed June 2005
ODPM (2004) The English indices of deprivation (revised), Office of the Deputy Prime Minister, ODPM Publications, London,
Openshaw, S. (1984) The modifiable areal unit problem. Concepts and Techniques in Modern Geography 38, Geo Books, Norwich

Author information

Dr Kenneth Field
Senior Lecturer in GIS
School of Earth Sciences and Geography
Kingston University London
Penrhyn Road
Kingston Upon Thames
Surrey KT1 2EE
kenneth.field@kingston.ac.uk