Socio-Ecological Mapping of Physical Activity Behaviours and Health Outcomes in Deprived Inner-City Communities

Geographical Information Systems Technical Report

Graham Smith and Rachel Davey
Staffordshire University,
Acknowledgements

The project is part of the National Prevention Research Initiative (NPRI), which is funded by a broad consortium of charities, research organisations and government, and is co-ordinated by the Medical Research Council Grant No. G0501287

We would like to thank the following people and organisations for their help with this project:

Christine Dover, for administrative support and help preparing data sets.
Rosemary Duncan, for the design and production of this publication.

Steve Smith (Assistant Chief Executive, Regeneration and Heritage) and John Nichol (Group Manager, Transport Planning) at Stoke-on-Trent City Council for allowing access to Transport Planning data sets.

Chris Jackson, Carol Birchall and Chris Salmon at Stoke-on-Trent City Council for providing Transport Planning data sets.

Chris Collins at the Knowledge Management Unit (Stoke-on-Trent City Council) for providing data sets.

Sgt Paul Rushton of the Staffordshire Police for arranging access to crime and anti-social behaviour data sets.

Denise Redfern, Mike Brunt and Lee Pointon at Stoke-on-Trent City council for providing crime and anti-social behaviour data sets.

Scott Grindey at RENEW North Staffordshire for providing data and guidance on regeneration zones and green space areas.

We would also like to thank the NPRI funding partners:
British Heart Foundation
Cancer Research UK
Department of Health
Diabetes UK
Economic and Social Research Council
Food Standards Agency
Medical Research Council
Research and Development Office for the Northern Ireland Health and Social Services
Chief Scientist Office, Scottish Executive Health Department
The Stroke Association
Wales Office of Research and Development, Welsh Assembly Government
World Cancer Research Fund

Authors
Graham Smith, Institute for Environment, Sustainability & Regeneration
Staffordshire University, Mellor Building, College Road, Stoke-on-Trent ST4 2DE, UK
Website: www.staffs.ac.uk/iesr     Email: G.R.Smith@staffs.ac.uk

Professor Rachel Davey, Centre for Sport and Exercise Research
Staffordshire University, Brindley Building, Leek Road, Stoke on Trent ST4 2DF
Website: www.staffs.ac.uk/schools/health/she/directory/rd/index
Contents

Acknowledgements 2

Contents 3

1 Introduction 7
1.1 Project Summary 7
1.2 Objective Measures of the Environment 7

2 Key concepts 9
2.1 GIS 9
2.2 Proximity (Distance) Measures 9

3 Study Area Selection 13
3.1 Lower Level Super Output Areas and the Index of Multiple Deprivation 2004 13
3.2 Study area selection process 14

4 Locating and classifying the residential population 18
4.1 Locating Residential Addresses 18
4.2 Measuring population 21

5 Locating Roads and Pathways 22
5.1 Aims 22
5.2 Data Sources 22
5.3 Methodology 24
5.4 Discussion 28

6 Measuring Street Connectivity 29
6.1 Aims 29
6.2 Methodology 29
6.3 Discussion 33

7 Proximity to Commercial Addresses and Food Outlets 35
7.1 Aims 35
7.2 Data sources 36
7.3 Methodology 36
7.4 Discussion 40
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 Locating environmental supports for physical activity</td>
<td>42</td>
</tr>
<tr>
<td>8.1 Locating physical activity facilities</td>
<td>42</td>
</tr>
<tr>
<td>8.2 Identifying areas of green space</td>
<td>45</td>
</tr>
<tr>
<td>8.3 Measuring access to environmental supports for physical activity</td>
<td>47</td>
</tr>
<tr>
<td>8.4 Discussion</td>
<td>49</td>
</tr>
<tr>
<td>9 Calculating land use mix</td>
<td>51</td>
</tr>
<tr>
<td>9.1 Aims</td>
<td>51</td>
</tr>
<tr>
<td>9.2 Data sources</td>
<td>51</td>
</tr>
<tr>
<td>9.3 Methods for calculating land use mix</td>
<td>52</td>
</tr>
<tr>
<td>9.4 Discussion</td>
<td>54</td>
</tr>
<tr>
<td>10 Mass Transport Provision</td>
<td>56</td>
</tr>
<tr>
<td>10.1 Aims</td>
<td>56</td>
</tr>
<tr>
<td>10.2 Data sources</td>
<td>56</td>
</tr>
<tr>
<td>10.3 Methodology</td>
<td>56</td>
</tr>
<tr>
<td>10.4 Discussion</td>
<td>58</td>
</tr>
<tr>
<td>11 Road Traffic and Street Safety</td>
<td>59</td>
</tr>
<tr>
<td>11.1 Aims</td>
<td>59</td>
</tr>
<tr>
<td>11.2 Road Traffic Levels</td>
<td>59</td>
</tr>
<tr>
<td>11.3 Accident Data</td>
<td>63</td>
</tr>
<tr>
<td>12 Crime Statistics</td>
<td>66</td>
</tr>
<tr>
<td>12.1 Aims</td>
<td>66</td>
</tr>
<tr>
<td>12.2 Data sources</td>
<td>66</td>
</tr>
<tr>
<td>12.3 Methodology</td>
<td>67</td>
</tr>
<tr>
<td>12.4 Discussion</td>
<td>67</td>
</tr>
<tr>
<td>References</td>
<td>69</td>
</tr>
<tr>
<td>Appendix 1 - Index of Deprivation Domains</td>
<td>71</td>
</tr>
<tr>
<td>Appendix 2 – OS MasterMap</td>
<td>74</td>
</tr>
</tbody>
</table>
List of Tables

Table 3.1  Deprivation quintiles of LSOAs involved in the selection process
Table 3.2  Final selection of LSOAs with IMD 2004 scores
Table 4.1  NLUD order UV070 residential groups
Table 6.1  Connectivity measures within 500m of two neighbourhoods
Table 7.1  Commercial addresses identified in the study area
Table 7.2  Density of retail addresses within 1km of each LSOA
Table 7.3  Percentage of residential addresses within x metres of a fast food outlet
Table 8.1  An example of green space accessibility results
Table 9.1  Land use classification categories
Table 9.2  HHI results for each study area
Table 9.3  Spatial Entropy results for each study area
Table 10.1  Access to a bus stop along the road and pathway network
Table 10.2  Bus stop density and average distance to a bus stop
Table 11.1  Road Traffic quintile classifications
Table 11.2  Road Traffic indices for LSOA study areas
Table 11.3  Accident casualty measures
## List of Figures

- **Figure 2.1** Euclidean buffer distance and network distance
- **Figure 2.2** Service Area calculated using network distance
- **Figure 3.1** Map of selected LSOAs and exclusion criteria within Stoke-on-Trent
- **Figure 5.1** An example of major roads removed from the pedestrian network
- **Figure 5.2** Identifying missing pedestrian routes
- **Figure 6.1** An example of a 4-way junction represented as two 3-way junctions
- **Figure 6.2** An example of ‘Effective Walking Area’ calculations
- **Figure 6.3** Comparison of street patterns within 500m of two neighbourhoods
- **Figure 8.1** An example of green space mapping
- **Figure 8.2** An example of mapped green space access points
- **Figure 8.3** An example of the green space accessibility analysis
- **Figure 11.1** An example of mapped road traffic classifications
- **Figure 11.2** Accident locations mapped with traffic levels
1 Introduction

This report outlines the datasets and methodologies required to create indices and measures of environmental determinants of physical activity behaviour using a Geographical Information System (GIS).

1.1 Project Summary

Increasing physical activity (PA) in the population is a priority for improved public health. This study provides a detailed mapping of the environment at lower super output area (LSOA) level in Stoke-on-Trent and the relationship between the environment, PA and health was evaluated. The environmental mapping aggregated data from available databases, augmented by local data gathering and validation, to produce a comprehensive geo-coded map of twelve LSOAs (covering a population ~18,000). Analysis, using geographical information systems (GIS) and associated software, was used to derive indices through which to evaluate the relationship between environmental characteristics and levels of physical activity and health, using hierarchical linear modelling. Environmental indices used included: proximity of PA spaces and facilities, street connectivity, land use mix, population density, mass transport provision, traffic, safety, crime, proximity of food outlets and shops/amenities, weather and indices of multiple deprivation. The mapping approach adopted a social ecological perspective, with a view to better understand (and, in future, change) the relationship between the environment and health behaviours, such as physical activity. The twelve neighbourhoods (LSOAs) were chosen (structured random sample) from the eligible LSOAs within Stoke-on-Trent. The community-based survey included a representative random sample from each LSOA from the Postcode Address File. Baseline data in the selected communities was collected by independent interviewer-administered survey covering physical activity, stages of change process (self-efficacy, intentions, attitudes, subjective norms) and outcomes, health, health-related quality of life, health behaviours, perceptions of the local neighbourhood and socio-demographic information (such as gender, age, ethnicity, socio-economic circumstances, education level, tenure). Data on health care utilisation and costs for PA participation was also gathered.

1.2 Objective Measures of the Environment

This technical report presents details of the GIS methodology used in the study and how the environmental indices were developed and the metrics used. Since this is one of the most detailed studies to have been undertaken in a UK context, the authors hope that this will be of interest to other researchers undertaking similar work in the UK.

This report describes how residential populations can be mapped and modelled at a
household level and, in addition, how a “walkable” network of roads and pathways can be produced. It also discusses the issues of proximity and accessibility and how these can be quantified in a GIS along with introducing other common GIS methodologies. The methodologies describe how proximity to various supports for physical activity behaviour was measured and how potential barriers to physical activity behaviour were modelled and quantified. It also describes how GIS was used as a tool to aid the selection of study sites.

For each environmental aspect that is described there is a discussion of the datasets used, the GIS methodologies implemented and the associated strengths and weaknesses. In addition, the report describes the specific metrics produced and how, potentially, they can be used.

The methods described produce a detailed mapped dataset of metrics that quantify potential environmental determinants of physical activity behaviour. These metrics have been used to highlight and quantify differences between study areas and to explore relationships between environmental factors and actual reported levels of physical activity behaviour (self reported and objectively measured) through the use of a large community survey (reported separately).
2 Key concepts

2.1 GIS

A Geographical Information System (or GIS) is software that provides a set of tools for the collection, display, manipulation and analysis of spatial (geographically referenced) data. GIS can be used to manipulate, analyse, and present information linked to geographical location and is being used increasingly in PA research. Spatial analytical methods using GIS have the potential to combine multiple datasets linked to methods of spatial analysis, to determine relationships between geographical patterns of disease distribution and social and physical environmental conditions.

The software package used by the research team for this particular project was ESRI ArcGIS – ArcInfo version 9.2. However, a number of GIS packages and tools exist and this report describes the general methodologies implemented using tools available with most GIS packages.

2.2 Proximity (Distance) Measures

The majority of metrics described in this report are derived using proximity measures; for example, the number of retail outlets within 200m of an area or the number of green space areas within 1km of individual households. There are two types of proximity measures used in this study:

2.2.1 Euclidean distance

Euclidean distance refers to the straight line distance or ‘as the crow flies’ distance. These distance measures are calculated in the GIS in two ways. The first and simplest method used is to measure the distance between two points. GIS software can be used to calculate the nearest point from one dataset (e.g. bus stop locations) to another.

The second method is to use a common GIS process known as buffer analysis. This method produces a defined buffer around a feature that is calculated using a specified distance and then features within the buffer can be identified. Buffer analysis is used in two ways in this study. It is used to identify features such as commercial addresses or PA facilities that are within 1km (or other specified distance) of the boundary of a study area. It is also used to identify features that are located within 1km (or other specified distance) of a point location such as a residential address.
2.2.2 Network distance

For measures concerned with physical access (or walking distance), such as distance to green space, a network analysis approach is preferred to the Euclidean approach.

Network analysis methodologies calculate distance from one object to another along a defined network of routes. In the case of measuring walking distances from a population to an environmental feature such as green space the network used may consist of roads and pathways available to pedestrians or cyclists. Figure 2.1 illustrates the difference between using a Euclidean distance of 500m around a residential address using a buffer and measuring all routes of 500m along a road and pathway network from the same address. Looking at the retail outlets and areas of green space shown in Figure 2.1 demonstrates that the number of accessible locations within 500m along the network is fewer than the number of sites inside the 500m buffer for that particular household. For example, nine retail outlets are within the 500m buffer but only two of these are accessible within 500m along the pedestrian network. Chapter 5 gives a detailed discussion of how this road and pathway network was created for the Stoke-on-Trent area and the datasets used in the process.

Figure 2.1 Euclidean buffer distance and network distance

In this study the tools provided by the Network Analyst extension of ArcGIS were used to perform distance analysis. The general process of network analysis is used to calculate a number of different types of measures. Firstly, it can be used to calculate the shortest route from an origin, such as a household address to a specified destination, such as a park. In addition to calculating the distance, the route itself can be mapped in the GIS.
Secondly, a closest facility analysis can be performed. This takes a series of origins, such as the residential addresses in a study area and for each one in turns identifies the closest facility (or a set number of facilities), such as sports facilities. In addition to identifying the closest facility it reports the distance and provides a mapped version of the shortest route.

Thirdly, a service area can be defined around a particular location at a set distance. This area is defined by joining the end points of all routes measured up to a defined distance away from an origin point. Figure 2.2 shows the 500m service area around a residential address and compares this with a 500m straight line buffer. In addition to the catchment boundary, a data layer containing all possible routes away from the address is produced. These routes are also shown in Figure 2.2

**Figure 2.2 Service Area calculated using network distance**

This method can be used to define a neighbourhood area around a location such as an address. This neighbourhood boundary can be used to produce ‘area based’ measures of the environment, such as the number of fast-food outlets within the service area.

Finally, the most commonly used analysis in this study has been the creation of an origin-destination matrix. This type of analysis takes a set of origins, such as the residential address locations in a study area, and calculates the distance along the network to all destinations, such as green space areas, up to a user defined distance, such as 1km. The resulting origin destination table can be interrogated to produce a number of measures.
2.2.3 Individual household measures and area-based results

The proximity measures produced in this study have been reported at a number of scales. For network analysis distances, individual measures have been produced for every individual household. These measures can be aggregated to provide area-based figures for a number of different geographical scales. Firstly, LSOA level figures have been produced for each area selected for the study. Secondly, other levels have been examined, such as, census output area and postcodes. For each level of aggregation a number of measures have been produced. For example:

- Average distance from residential addresses to the nearest commercial premises (or food outlet) by type (also minimum, maximum and range of distances).
- Population (and percentage of total) within 200m, 400m...1km of commercial premises (or food outlets) by type.
- Population (and percentage of total) within x metres of 1, 2, 3... commercial premises (or food outlets) by type.

In addition to household accessibility measures, each LSOA study area boundary has been used to produce a number of area-based figures, such as the areal density of retail outlets. Further to this the boundary of each LSOA has been buffered using distances from 200m up to 1km. The same density measures have been calculated within each buffer area. This provides a useful comparison to accessibility measures and captures the variety of commercial premises.

When assessing each measure it is important that the geographical area (buffer distance) is small enough to capture sufficient local variation and displays the full range of data. For some measures, as the buffer size increases the area becomes too large to capture the detail of the interaction between individuals and their immediate environment.
3 Study Area Selection

3.1 Lower Level Super Output Areas and the Index of Multiple Deprivation 2004

Super Output Areas are available at three levels and the geographical unit selected for this study was Lower Level Super Output Area (LSOA). These spatial units were chosen because they provide an approximate neighbourhood level of analysis due to their population size. In England there are 32,482 LSOAs and they are defined by the Office of National Statistics (ONS) as:

“Minimum population 1000; mean 1500. Built from groups of Output Areas (typically 4 to 6) and constrained by the boundaries of the Standard Table (ST) wards used for 2001 Census outputs.” (ONS 2005)

Stoke-on-Trent has 160 LSOAs. The aim of the study was to look at ten deprived LSOAs and two more affluent LSOAs within the city to ensure that a representative range of deprivation was included. In order to do this the Index of Multiple Deprivation 2004 (IMD) was used as an indicator. In Stoke-on-Trent, 50.6% of the LSOAs are ranked in the 20% most deprived nationally. Only 8.8% of LSOAs are ranked in the 40% least deprived nationally.

The index is based on 37 indicators which are combined to make seven domains of deprivation. These are:

- Income deprivation
- Employment deprivation
- Health deprivation and disability
- Education, skills and training deprivation
- Barriers to housing and services
- Crime
- Living environment

Further information regarding the components of the Index of Multiple Deprivation 2004 for England can be found in Appendix 1.

For each LSOA, a score is produced for each indicator and then each domain. Individual domain scores are then weighted and summed to create the overall IMD score for the LSOA. This IMD score forms the basis for a final ranking of LSOAs.

It is important to note that the IMD is a relative ranking of deprivation. Using this dataset population weighted deciles (and corresponding quintiles) were created for England. In
order to create the deciles the rank was used to place each of the LSOAs into deciles of equal population. Deciles of equal population were preferred to those of equal LSOA count as the analysis gives a population-based distribution. Furthermore classifying all the LSOAs depending on their national ranking in the IMD allowed us to make comparison with some other studies.

In all cases decile 1 is the most deprived and decile 10 is the least deprived. It is important to understand what these deciles represent. Essentially decile 1 has the largest concentration of deprived people while decile 10 has the smallest concentration of deprived people. Decile 1 is not ‘the poorest 10% of the population’ as some of the poorest people will live within pockets within less deprived LSOAs, nor is it ‘the 10% most deprived LSOAs’ as a population weighting has been applied.

It is also important to realise that the population within a LSOA and within a decile will vary in their characteristics. The IMD is providing a measure for a group of people, not a precise measure for every individual. Within area-based studies this is a well-known limitation, known as the ecological fallacy, which requires a caveat to be placed on any area-based analysis. However Spicker (2001) has disputed the issue of ecological fallacy when dealing with deprivation. One of the key arguments he makes is that poor areas are more than just the sum of experience of poor individuals. More wealthy people in deprived areas are also affected through a lack of resources in an area.

In addition to the IMD data, a number of census variables were collected at LSOA level (and nested census output areas) in order to create a profile of each area. These included: age structure and gender, household type, tenure, migration.

3.2 Study area selection process

Due to the nature of the study the selection of the 12 LSOAs could not be a completely random exercise as a range of deprivation was needed. Furthermore, many parts of Stoke-on-Trent are experiencing some level of urban renewal. As a product of various regeneration and renewal initiatives the urban landscape in these areas can be subject to rapid change. Also, these areas can experience higher than average population churn rates. Due to the duration of this study it was necessary to select LSOAs that did not fall inside or within 200m of the boundaries of any of these renewal areas. These areas were identified using digital boundaries provided by RENEW North Staffordshire.

The result of applying these selection criteria left 25 LSOAs out of 160 in Stoke-on-Trent. The final chosen LSOAs needed to represent the range of deprivation within Stoke-on-Trent with a reasonably representative proportion coming from each deprivation quintile. Therefore the next stage involved the random selection of 10 LSOAs (out of 22) from the most deprived 3 quintiles and 2 LSOAs (out of 3) from the 2 least deprived quintiles.
A summary of the deprivation profile of the LSOAs at each stage of selection can be seen in Table 3.1.

**Table 3.1 Deprivation quintiles of LSOAs involved in the selection process**

<table>
<thead>
<tr>
<th>IMD National Ranking Quintile</th>
<th>Stoke-on-Trent LSOAs</th>
<th>Stoke-on-Trent LSOAs</th>
<th>Stoke-on-Trent LSOAs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>%</td>
<td>Count</td>
</tr>
<tr>
<td>1 (Worst 20% LSOAs)</td>
<td>81</td>
<td>50.6</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>26.3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>14.4</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>7.5</td>
<td>2</td>
</tr>
<tr>
<td>5 (Best 20% LSOAs)</td>
<td>2</td>
<td>1.3</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>160</td>
<td>100</td>
<td>25</td>
</tr>
</tbody>
</table>

The final criterion for area selection was that wherever possible the selected LSOAs should not be adjacent (to minimise contamination) as part of the research is to deliver a physical activity “intervention” in 6 of the LSOAs and with 6 LSOAs as “controls”. In order to test this, each of the randomly chosen areas was mapped in a GIS. If two areas were found to be adjacent then one of the areas was removed and replaced by another from the same quintile. The newly selected LSOAs were then mapped and if also found to be adjacent to another LSOA then the process was repeated. The final list of selected LSOAs can be found in Table 3.2

**Table 3.2 Final selection of LSOAs with IMD 2004 scores**

<table>
<thead>
<tr>
<th>LSOA Code</th>
<th>Area Name</th>
<th>IMD Score</th>
<th>IMD Rank</th>
<th>Decile</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01014269</td>
<td>Sneyd Green</td>
<td>55.73</td>
<td>1,401</td>
<td>1</td>
</tr>
<tr>
<td>E01014300</td>
<td>Sandford Hill - East</td>
<td>46.31</td>
<td>3,002</td>
<td>1</td>
</tr>
<tr>
<td>E01014345</td>
<td>Stoke</td>
<td>43.11</td>
<td>3,723</td>
<td>2</td>
</tr>
<tr>
<td>E01014279</td>
<td>Heron Cross</td>
<td>38.71</td>
<td>5,012</td>
<td>2</td>
</tr>
<tr>
<td>E01014353</td>
<td>Hanford</td>
<td>32.63</td>
<td>7,097</td>
<td>3</td>
</tr>
<tr>
<td>E01014278</td>
<td>Sandford Hill - West</td>
<td>32.22</td>
<td>7,244</td>
<td>3</td>
</tr>
<tr>
<td>E01014338</td>
<td>Trent Vale</td>
<td>21.33</td>
<td>12,855</td>
<td>4</td>
</tr>
<tr>
<td>E01014297</td>
<td>Adderley Green</td>
<td>20.14</td>
<td>13,769</td>
<td>5</td>
</tr>
<tr>
<td>E01014292</td>
<td>Hartshill and Penkhull</td>
<td>16.66</td>
<td>16,544</td>
<td>6</td>
</tr>
<tr>
<td>E01014303</td>
<td>Meir Hay</td>
<td>15.95</td>
<td>17,226</td>
<td>6</td>
</tr>
<tr>
<td>E01014265</td>
<td>Baddeley Green and Milton</td>
<td>8.88</td>
<td>25,289</td>
<td>8</td>
</tr>
<tr>
<td>E01014350</td>
<td>Trentham</td>
<td>8.16</td>
<td>26,247</td>
<td>9</td>
</tr>
</tbody>
</table>
The map in Figure 3.1 shows the final selected areas and the location of the exclusion criteria, such as the renewal areas.

The first thing that this map illustrates is that a number of areas are within close proximity of renewal areas. An ideal scenario would be that all study areas are at distance away from renewal areas that represents a considerable barrier to any population from the study area having easy walkable access to renewal areas. For the purposes of this study this would be a distance of approximately 800m to 1km. However, given that Stoke-on-Trent is only approximately 8km (West to East) 16km (North to South), this is difficult to achieve.

The second issue that is illustrated in Figure 3.1 is that there are a number of study areas in relatively close proximity. Once again the ideal scenario would be to make sure that areas are not easily accessible from one another. This is a particularly important issue if areas are split into control and intervention areas for a particular study.

Although these issues needed to be acknowledged for this particular study they do not prevent the results produced from being valid. The methodologies used in this report generally classify the study areas themselves or the population’s level of access to various environmental features. The result of this is that study areas found within close proximity will display some similarity for measures involving access because they share part of the same environment.
Figure 3.1 Map of selected LSOAs and exclusion criteria within Stoke-on-Trent

Selected LSOAs
IMD National Quintile
- Top 20% (most deprived)
- 20% to 40%
- 40% to 60%
- 60% to 80%
- Bottom 20% (least deprived)

Urban Renewal Areas
- Area Boundary
- 200m buffer
4 Locating and classifying the residential population

4.1 Locating Residential Addresses

4.1.1 AddressLayer 2

Residential addresses need to be located as precisely as possible in order to make distance calculations as accurate as possible. A number of address and postcode products exist but the majority are all built from the same source data combining data from Royal Mail, Ordnance Survey and the Office of National Statistics (ONS). The dataset chosen to locate and classify residential address locations was ‘Ordnance Survey MasterMap AddressLayer 2’. This dataset gives the geographic point location (easting and northing) of each postal address within the UK to the nearest 1 metre.

AddressLayer 2 has a multiple occupancy count field. This count is supplied by Royal Mail and records the number of residences that are behind a single delivery point. Instead of classifying the addresses as a single residence the multiple occupancy count can be used as a population multiplier.

Residential housing

The positional accuracy of coordinates is reported in AddressLayer 2 (Positional Quality and Accuracy of Position fields). In the 12 study areas there were 8,198 residential addresses and of these 8,167 (99.6%) were in a final surveyed position. The remaining 31 addresses were located at the postcode unit mean which means that they will be very close to their actual location, i.e. within the correct street.

In addition each address has been classified by Ordnance Survey into a Base classification which gives an indication of the type or use of an address. This Base classification has been grouped into two commonly used classifications from the National Land Use Database (NLUD) and the Valuation Office Agency (primary description field and special category codes).
The NLUD classification can be used to identify residential addresses. Table 4.1 lists the residential populations that are present in this classification. The focus of this study is on the usual resident population of each study area and therefore the U071 – Dwellings category was used to identify all residential addresses.

Table 4.1 NLUD order UV070 residential group

<table>
<thead>
<tr>
<th>Group</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U071</td>
<td>Dwellings</td>
<td>Houses and flats for individuals and families living as a single household, including adjoining garages, gardens, non-thoroughfare service and distribution roads and pathways. + Caravan sites and mobile homes used as permanent dwellings. + Sheltered residential accommodation with separate front entrances.</td>
</tr>
<tr>
<td>U072</td>
<td>Hotels, boarding and guest houses</td>
<td>Hotels, B&amp;B’s, boarding houses, and residential clubs (where no significant element of care is provided).</td>
</tr>
<tr>
<td>U073</td>
<td>Residential institutions</td>
<td>Residential accommodation for provision of care e.g. old peoples’ homes, children’s homes and other non-medical homes. + Residential schools and colleges and training centres, including university and hospital residences. + Communal residences e.g. barracks, monasteries and convents.</td>
</tr>
</tbody>
</table>

Source: OS AddressLayer 2 Technical Specification document

A further description of Address Layer 2 can be found in Appendix 2.

4.1.2 Methodology

The methodology for identifying residential address locations using AddressLayer 2 followed the following procedures:

- The first stage of the process is to import OS MasterMap AddressLayer 2 into a GIS. AddressLayer 2 can be provided in csv (comma separated values) text files which will be easier to import than other layers of OS MasterMap which are delivered in a compressed gml format (geography markup language).
• The second stage is to use fields contained in the data to remove PO-Box addresses and demolished properties. PO-Box addresses are commercial addresses that are given the geographic coordinates of the nearest royal mail depot.

• The third stage is to identify residential addresses by selecting addresses with the U071 – Dwellings NLUD category.

• The fourth stage is to identify residential addresses that are multi-occupancy locations. This can be done using the multiple-occupancy count field present in AddressLayer 2. AddressLayer 2 also includes features called Multiple-Occupancy Without Post Address (MOWPA) which are point locations representing each address behind a delivery point. At locations where these features are present it is not necessary to use the multiple-occupancy count as a multiplier of population.

• The final step is to identify addresses that have temporary coordinates. The level of accuracy of temporary coordinates can also be noted. If a large number of features in an area had temporary coordinates (such as new build areas) then these locations could give unreliable results. In this particular study it has already been noted that only a very small number (0.4%) of addresses have provisional coordinates.

The steps above could largely be completed using other address datasets such as OS MasterMap AddressLayer and AddressPoint. However these datasets do not include the MOWPA features and AddressPoint does not include a multiple-occupancy count field. If address data sets are not available then postcode products such as OS CodePoint could be used to map postcode centroids and the residential address count fields could be used to help multiply populations. Alternatively, individual local authorities could use their Local Land and Property Gazetteer (LLPG) to locate residential addresses.
4.2 Measuring population

4.2.1 Assigning a population to each residential address

Population data at a household level does not exist. Therefore area-based population figures have been used to approximate household populations. The figures used were taken from the IMD 2004 because this dataset was used to help select the study areas. The IMD population figures represent mid year 2001. The methodology used is discussed in the reference report for the IMD 2004 (See Appendix 1).

The population of each LSOA study area was divided by the number of residential locations identified. The residential locations are the identified residential addresses plus the addition of the multiple occupancy figures where required. This provided an average household size that can be applied to each of the residential locations. This method assumes the same level of occupation per household because data on household level populations does not exist. Many studies use a similar approach when household level data is not available and population near to particular facilities or environmental features is being estimated (Walker, 2003).

4.2.2 Calculating Population Density

The population density of each study area was calculated by simply dividing the population by the area (in km$^2$). Some studies in the USA (Frank et al., 2005) and Australia (Leslie et al., 2005) have measured the population density within areas of residential land use. This has not been replicated here due to differences in the planning system; primarily there is greater mixed use in the UK and it would be very difficult to measure the density in only residential areas. Measures of land use mix are discussed fully in Chapter 9.
5 Locating Roads and Pathways

5.1 Aims

The aim of this chapter is to describe the datasets and processes used to create a detailed GIS layer of roads and pathways.

An accurate representation of roads and pathways is necessary to successfully produce a number of the metrics described in this report. These metrics include: accessibility measures, utilising network analysis methodologies and metrics requiring the length of road to be calculated.

5.2 Data Sources

5.2.1 Roads

A digital mapped road network was taken from Ordnance Survey’s MasterMap Integrated Transport Network (ITN) layer. The ITN represents all road centrelines to 1 metre accuracy, thus making it suitable for applications such as measuring accurately the true length of a road and producing accurate accessibility measures using methods of network analysis as described in section 2.2.2.

The ITN layer provides a descriptive term field that contains a classification of roads. This classification contains motorways, primary A-roads, trunk A-roads, A-roads, B-Roads, minor roads and local streets as the major categories of roads.

The ITN also contains a field detailing the ‘nature’ of a road feature. This field indicates whether the road is a single or dual carriageway. Single carriageway roads are represented as a single line and dual carriageway roads as 2 lines, one for each side of the road. The nature field also identifies slip roads and roundabouts.
The ITN contains a large amount of road routing information. The majority of these fields relate to traffic restrictions such as a no entry or one way street. The data does include information on bridges which, along with slip roads and roundabouts, can be highlighted and treated as a special case.

This dataset was used for methodologies that require road network statistics, e.g. length of road by traffic levels.

5.2.2 Pathways, cycle paths and alleyways

GIS data showing the location of cycle routes in the city was obtained from Stoke-on-Trent City Council. This data had a number of uses. Firstly, it was used to locate some pathways that linked roads and other pathways together. Secondly, it was used to indicate stretches of road that had cycle paths or that were part of a designated cycle route.

It was not possible to obtain an existing data set to locate all pathways and interconnecting alleyways within Stoke-on-Trent. At the time of writing, no data is held by the city council that digitally maps pathways. One potential solution was to utilise data held in the Topography layer of OS MasterMap. However, this solution was not pursued due to the prohibitive cost of the data. If this data had been available it could only have been used to identify land areas that are classified as pathways; it would not provide a centreline of all pathways.

The solution used here to identify pathways was to utilise aerial photography mapped in a GIS to pick out clear, non-roadside pathways that connected to existing roads data and to each other. In addition to usual pathways, established routes across areas of open space were also identified. Once located on the photography the features where traced onscreen

Cycle Path

Cycle path closed
to digitise the pathways as a GIS layer. In addition to the use of the aerial photography the study areas were visited by the research team and any other interconnecting pathways and alleyways were noted. These areas were then found in the GIS and mapped onscreen using the aerial photography. This was a very time consuming process, but it resulted in a high quality representation of pathways which was needed to accurately measure accessibility.

An alternative method would have been to survey pathways in the field using GPS equipment. However, given the size of the study areas plus surrounding areas up to 1 kilometre away, it would have taken an unacceptable amount of time to survey. Also, once the data had been collected it would need to be cleaned and edited in the GIS for use with the roads data.

5.3 Methodology

5.3.1 Producing a pedestrian road dataset

This section describes how the road and pathway datasets were combined to create one definitive pedestrian layer. This final layer represented all walkable roads and pathways.

The ITN roads data is not designed to be representative of a pedestrian network; it is strictly the centre lines of all roads. However, the location of the road centre lines provides a good proxy for the routes taken by pedestrians along the roadside (unlike some areas of the USA the vast majority of urban roads have pavements on both sides in the UK). However, it would not have been sensible to use all of the roads as a proxy of pedestrian routes because a number of them would not be walkable.

Non-walkable roads were defined as routes that would be difficult or dangerous for pedestrians to use. The ITN data does not supply information on pavements and therefore a number of assumptions had to be made. Non-walkable roads were defined as:

a. Dual carriageways (major A roads) that have no pavements.
b. Slip roads to dual carriageways
c. Road bridges over major roads that were known not to have pavements for pedestrians.

Major roads were selected using the descriptive term field to select motorways, primary A-roads, trunk A-roads. In addition to these roads, the nature field was used to select dual carriageway roads. These roads were removed from the study areas if they were known not to have a pedestrian friendly route such as a pavement or crossing. Therefore, local knowledge of the areas needed to be obtained. This was possible because of the small number of major roads in the study areas. Figure 5.1 shows an example of two major dual carriageways that were removed from the study area.
In the absence of local knowledge the following roads could be assumed to be unsuitable for pedestrians and therefore be removed:

- Motorways
- All slip roads
- All A-roads that are connected to slip roads

It would be inappropriate to remove all dual carriageways from the data. It is likely that a number of these, particularly in built up areas, will have pavements or neighbouring routes for pedestrians. Therefore these types of roads should be located and further judgements made about their possible inclusion in the dataset. In this study these roads did not have slip roads. These roads illustrate the added issue of crossing points on major roads. ITN data does not provide the location of pedestrian crossings.

Due to the fact that the ITN is a road network it had to be adapted for pedestrian use by manipulating parts of the dataset when dealing with junctions. This is because pedestrians may be able to move from one part of the network with freedom whereas vehicles have restricted movements.

Once a decision was made, and certain roads were removed, this left an approximated pedestrian road network. Although a number of assumptions have been applied to certain roads, this dataset provides a good spatial representation of the routes available to pedestrians along the road network.
5.3.2 Preparing road and pathway data for network analysis

In order to utilise network analysis methodologies, as described in Chapter 2, a merged and ‘connected’ layer of pedestrian road routes and digitised pathways needs to be created.

**Merge data**
Firstly the two datasets, the pedestrian road layer and the digitised pathways, need to be combined using a standard GIS tool to ‘merge’ them together. The ‘merge’ tool not only produces a single spatial layer, but also preserves the attributes applied to each layer.

**Connect line features**
Once a merged dataset has been created the next step is to use GIS editing tools to ensure that all lines stored in the layer are connected. For example, a digitised pathway linking two roads will be represented by a single line. This line needs to be ‘connected’ to each road in order for the network analysis tools to recognise the pathway as a valid route between the two roads. This can be performed using ‘extend’ tools and the ‘snapping’ tools. The extend tools allow line features to be extended until they reach the edge or end of another feature. Snapping tools can be used in a similar way to ‘snap’ the end of a feature to the edge or end of a neighbouring feature.

**Splitting stretches of road**
The next stage required is to make sure that each line segment is split to make two lines where it meets another line at a junction. It is important to note that this is already the case in the ITN roads data and therefore only junctions involving newly added pathways need to be altered. This is done using the ‘split’ editing tools. This process, along with the previous process can be automated using a ‘snap and split’ tool that can be downloaded from the ESRI – ArcScripts website (http://arcscripts.esri.com).

**Creating junction nodes**
Once lines have been merged, snapped and split, the GIS can be used to create a point, called a node, at each junction between line features and at the end of a line feature that represents a dead-end. This was done using the Fnode-Tnode script downloaded from the ESRI – ArcScripts website. This produces a point layer containing one point per every junction and every end point in the road and pathway layer. In addition to the location of the junctions, the number of lines that converge at each junction is recorded as an attribute. This attribute will help to identify junctions of a specific type, as well as cul-de-sacs in Chapter 6, for the creation of street connectivity measures.

An additional benefit of creating these nodes is that it highlights connectivity errors in the road and pathway data if a node is not found at a supposed junction.

**Checking for missing pedestrian routes**
Examination of the roads and pathways at this stage of the process revealed that there
were occasionally dead ends in close proximity to one another and it was deemed necessary to pick these out and examine whether they should actually be connected because there is a direct pedestrian route between them.

An automated process of identifying these locations was developed. Every dead end was identified by selecting nodes or junctions that had a line count of 1.

Next, a buffer of 15 metres was applied around each of the selected nodes. The GIS was used to merge buffered areas that overlap to form one polygon or area instead of having one area for each point. If two buffers overlapped, it meant that the two nodes (dead-ends) were less than 30 metres apart.

Then a point in polygon analysis was used to count the number of nodes that fell within each of the buffer areas. Any buffer that had a count of 1 node was considered to definitely be a dead end. Whereas, any buffer that had a count greater than 1 was highlighted as being an area where there was potentially a direct path between two or more points. Figure 5.2 gives an example of each of these scenarios.

![Figure 5.2 Identifying missing pedestrian routes](image)

Aerial photography was then used to examine whether there was a direct link between the two (or more) points. If this was deemed to be the case, then a new pathway was added to the road and pathway data to create the link between the nodes.

After the new links had been identified and added to the data the Fnode-Tnode script was re-run to create a new set of nodes.
Building a network dataset
Once the road and pathways data has been prepared it can be used to create a ‘network dataset.’ This is done using the ‘Create network dataset’ tool provided in ArcGIS Network Analyst. The tool allows the measuring units of the network to be specified (this study used metres) as well as any restrictions of movement through the network. This study did not attach any restrictions to movement through the network because it focuses on pedestrian movement. Restrictions are generally used for traffic routing analysis, such as one-way streets.

It is important to note that if the processes above are not completed then the network dataset will not function properly. All lines must be split at junctions and all lines must be connected at junctions.

5.4 Discussion
The approach used here (particularly as there was no access to pathway data) has produced a very detailed representation of a pedestrian network throughout the study areas.

No road and pathway network data set can be completely accurate because routes are identified using road centrelines meaning that the network may over-estimate distance measures. This is because a centreline does not allow for cutting corners at junctions, particularly over accessible open space such as a car park. However, this is a small and unavoidable limitation of this type of analysis and compares with approaches used in other PA studies (Dill, 2003 and Frank et al., 2005).

A large number of GIS resources are available to achieve the same results as the tools and processes used here. As long as the output of the tools used produces a network of roads and pathways as described above then the outputs will be effectively the same.
6 Measuring Street Connectivity

6.1 Aims

Street connectivity measures quantify how well connected a street network is. Areas dominated by a series of cul-de-sacs have lower levels of connectivity than areas containing a high proportion of junctions. The more ‘connected’ a neighbourhood, the easier it is for a resident to find a direct route to a destination, such as a park or a shop. Routes were calculated using roads and pathways that are accessible by active transport; pedestrians and cyclists.

A number of connectivity measures have been developed and reviewed in other studies in the USA (Dill, 2003). This chapter identifies six methods that can be used successfully in a UK setting.

6.2 Methodology

The methods utilise the pedestrian roads and pathways and junctions data produced in Chapter 5. The indices calculated use the counts and types of junctions and the counts and lengths of road segments.

6.2.1 Classifying junctions

Identifying the type of junction

In order to use the methods described in the following sections the junctions in an area need to be classified into 1-way (dead ends or cul-de-sacs), 2-way, 3-way 4-way etc. Due to the nature of the ITN roads data it is not possible to assume that every junction should be classified by simply looking at the number of unique road segments that converge at each point. Digitising of roads and pathways within a GIS often result in road centrelines on either side of a junction being offset by a few metres and this can cause the number of junctions at a location to be misrepresented. Figure 6.1 gives the example of two 3-way junctions that are in close proximity within the pathway data and therefore, in reality, actually function as one 4-way junction.
A solution was required to manipulate the current junctions to better reflect the number of actual junctions and their associated function (e.g. 3-way, 4-way). A solution to these problems has been developed in the Twin Cities Walking Study in the USA (Forsyth, 2007). This was achieved by applying a buffer around junctions of 10 metres and where buffers overlapped (because they were within 20 metres of each other) the centroid of the combined buffer areas was used to create a new junction. Then a number of rules were applied to use the total count of junction points and road segments within the buffer area to reclassify the ‘type’ of the new junction. In the example shown in Figure 6.1 the total number of junctions within the dissolved buffer area is 2 and the total intersection count is 6 (3-way + 3-way). Therefore all dissolved buffer areas matching these counts are classified as a 4-way junction.

Removing junctions from the end of cul-de-sacs
Junctions classified as 3-way (and occasionally 4-way) can often be observed near to the dead end of a cul-de-sac. This is because very small segments of road are recorded, thus giving the indication that there is a junction where just a turning point at the end of the road exists. These junctions have been removed from any calculations because they produce an over-estimation of the level of connectivity in an area.

Remove junctions involving major (non-pedestrian) roads
It is important to note that this analysis uses the roads and pathways produced using the methodology described in Chapter 5. This data set is representative of the pedestrian road and pathway network. Therefore, no intersections involving major roads or slip roads have
been included in any calculations.

6.2.2 Road and pathway length per unit area

The length of roads and pathways per unit area was calculated to give a measure of the density of the pedestrian network. It is assumed that areas with a higher density of roads and pathways will be better connected. This was calculated by dividing the total length of roads and pathways in an area by the land area (in sq. kilometres).

The total length of roads classified as dual-carriageways was divided by two because they are represented using two centrelines and therefore the total length is being double counted.

This metric was calculated within every study area, and also within a buffer area of 500m from each study area.

6.2.3 Intersections (3-way or more) per unit area

Intersections of three or more road segments provide multiple route choices for pedestrians. It is assumed that the higher the density of intersections, the higher the level of connectivity in an area. This metric was calculated by dividing the number of 3-way (or more) intersections in an area by the land area (in sq. kilometres).

Intersections with 3 or more contributing roads and pathways were calculated using the junctions created in section 5.3.2 and the methodology in section 6.2.1.

This metric was calculated within every study area, and also within a buffer area of 500m from each study area. In addition, this metric has been calculated within a buffer area of 500m from every residential address location (identified in Chapter 4).

6.2.4 Ratio of 3-way (or more) intersections to all intersections

It is assumed that the higher the proportion of 3-way (or more) intersections, the higher the level of connectivity in an area. This metric was calculated by dividing the number of intersections classified as 3-way or more by the total number of intersections in an area (including cul-de-sacs and dead ends). Values range from 0 to 1 and areas with values greater than 0.7 are considered favourable (Smart Growth Index).

Intersections were classified using the methodology in section 6.2.1.

This metric was calculated within every study area, and also within a buffer area of 500m from each study area. In addition, this metric has been calculated within a buffer area of 500m from every residential address location (identified in Chapter 4).
6.2.5 Gamma index

The gamma index provides a measure of the level of connectivity in a network. It provides a ratio of the number of links present in a network to the maximum possible number of links in a network. The values range from 0 to 1, with 1 representing a network with the maximum number of links present. These values can be reported as a percentage of connectivity (Dill, 2003), e.g. a gamma index of 0.54 represents a network that is 54% connected. The gamma index is calculated as follows:

\[ \text{Gamma Index} = \frac{L}{L_{\text{max}}} \]

\[ L_{\text{max}} = L/3(V - 2) \]

where, \( L \) is the number of links (road and pathway segments) in a network and \( V \) is the number of nodes (intersections and dead ends).

6.2.6 Alpha index

The alpha index is a measure of circuitry in a network. The alpha index calculates the ratio of the number of circuits present in a network to the maximum possible number of circuits. The values range from 0 to 1, with 1 representing a network with the maximum number of circuits present. Therefore, areas with a higher alpha index are considered to have higher connectivity. The formula for calculating the alpha index is as follows:

\[ \text{Alpha Index} = \frac{(L - V) + 1}{2V - 5} \]

where, \( L \) is the number of links (road and pathway segments) in a network and \( V \) is the number of nodes (intersections and dead ends).

6.2.7 Effective walking area ratio

The effective walking area (Dill, 2003) is a measure of how connected the street network is. This measure is a ratio of area within \( X \) pedestrian network distance to area within \( X \) distance radius.

The area within \( X \) distance radius can be measured by applying a buffer around a point or area as described in section 2.2.1. and calculating the area. The area within \( X \) network distance is calculated by using network analysis tools to produce a service area. This is described in section 2.2.2. The service area produced gives an approximate measure of the area that can be reached by a pedestrian. Figure 6.2 illustrates these calculations for 2 areas.
This study has measured the ratio of area within 500m walking distance to area within 500m radius of every residential address location (identified in Chapter 4). In addition the ratio could be measured from each study area centroid.

6.3 Discussion

In the UK, although street patterns often seem well connected, the indices produced here show that a marked variation can be found between areas. Areas dominated by a series of cul-de-sacs will display lower levels of connectivity than areas containing a high proportion of junctions, such as, traditional areas of terraced-housing. Figure 6.3 shows the street pattern around two contrasting study areas. The Stoke LSOA is made up of a large number of terraced streets whereas the Hanford LSOA is an area of detached housing built around a series of cul-de-sacs. Table 6.1 shows the values of the six connectivity measures calculated for each area. These measures consistently report a greater level of connectivity in the Stoke LSOA.
Figure 6.3 Comparison of street patterns within 500m of two neighbourhoods

Table 6.1 Connectivity measures within 500m of two neighbourhoods

<table>
<thead>
<tr>
<th>Type of measure</th>
<th>Stoke</th>
<th>Hanford</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of roads and pathways per sq.km</td>
<td>18.85</td>
<td>7.76</td>
</tr>
<tr>
<td>Alpha Index</td>
<td>0.18</td>
<td>0.11</td>
</tr>
<tr>
<td>Gamma Index</td>
<td>0.46</td>
<td>0.41</td>
</tr>
<tr>
<td>Density of 3-way junctions per sq.km</td>
<td>147.26</td>
<td>38.15</td>
</tr>
<tr>
<td>Proportion of junctions 3-way or more</td>
<td>0.83</td>
<td>0.66</td>
</tr>
<tr>
<td>Average effective walking area ratio per residential address</td>
<td>41.37</td>
<td>37.81</td>
</tr>
</tbody>
</table>

The six methods of calculating street connectivity described above were chosen because they produce metrics that are easy to understand and can be compared to other studies.

Many more measures have been developed and reviewed in other studies in the USA (Dill, 2003). However, many of the measures are more appropriate in a USA urban setting and not appropriate for the UK. One example, ‘average block size’, has often been used as a crude measure of connectivity. However, a ‘block-based’ street pattern is rarely seen in UK urban areas.
7 Proximity to Commercial Addresses and Food Outlets

7.1 Aims

This chapter focuses on producing measures of the proximity and density of commercial addresses (or land use) with food outlets picked out as a notable sub-group.

It is believed that neighbourhoods within close proximity to a wide range of local shops and services, particularly food outlets, will be more likely to observe a higher number of ‘walking’ trips by its residents. Therefore the density or proximity of commercial properties could have a relationship with the general level of PA within an area.

In terms of proximity, the measures focused on retail outlets, local services, such as banks and post offices and on food outlets. These are the types of commercial locations that are more likely to receive regular visits by local populations.

The type of food outlet is examined in further detail because the level of choice within an area is seen as having potential impacts on the health of a neighbourhood. Recent studies in the USA present evidence that areas with greater density of fast food outlets have a higher prevalence of obesity and physical inactivity (Reidpath et al., 2002).

The first methodology used produced a measure of the density of broad categories of commercial use and food outlets for each study area. The second methodology produced a measure of the distance from each household in a study area to each ‘walkable’ commercial destination.

Local Shops and facilities
7.2 Data sources

This method uses the residential address locations identified in Chapter 4 and the roads and pathways identified in Chapter 5.

7.2.1 OS MasterMap – AddressLayer 2

The AddressLayer 2 dataset was used to identify, locate and classify commercial addresses for use within the study. This dataset gives the geographic point location (easting and northing) of each postal address within the UK to the nearest 1 metre.

In addition each address has been classified by Ordnance Survey into a Base classification which gives an indication of the type or use of an address. This Base classification has been grouped into two commonly used classifications from the National Land Use Database (NLUD) and the Valuation Office Agency (primary description field and special category codes).

More detailed information regarding Address Layer 2 can be found in Appendix 2.

7.3 Methodology

7.3.1 Identify and classify commercial addresses

The first step in this method was to identify all non-residential addresses within Address Layer 2 as ‘potentially’ commercial addresses. The NLUD and Valuation Office classifications were then used to group addresses into 6 broad categories. These categories can be seen in Table 7.1. The ‘general commercial’ group, which makes up 60% of the addresses found to be within the study areas, represents all commercial services and any addresses that were not easily identifiable as falling into one of the other groups.

A local services category was created as a sub-group of general commercial. This group represents services that will be used on a regular basis by the local community, such as banks, post offices, laundrettes, hairdressers etc. and was identified using the Valuation Office classification.

Food outlets were also identified using the Valuation Office classification. The food outlets are made up of the following categories:
- Eating and drinking category.
- Food made for consumption off premises (cafes and fast food).
- Food for consumption on premises (restaurants and public houses).
- A subset of the retail category. Addresses classified as food retail using the following subgroups: convenience stores, off licences, specialist (such as butchers, bakers etc.) and supermarkets.
<table>
<thead>
<tr>
<th>Group</th>
<th>Total</th>
<th>Distance from Study LSOAs</th>
<th>Inside</th>
<th>400m</th>
<th>600m</th>
<th>1km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>16</td>
<td>Inside</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating and Drinking</td>
<td>239</td>
<td>Inside</td>
<td>33</td>
<td>105</td>
<td>47</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>7</td>
<td>43</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Entertainment</td>
<td>33</td>
<td>Inside</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>14</td>
<td></td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>General Commercial</td>
<td>1,883</td>
<td>Inside</td>
<td>298</td>
<td>769</td>
<td>378</td>
<td>438</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>99</td>
<td></td>
<td>37</td>
<td>58</td>
</tr>
<tr>
<td>Industrial</td>
<td>238</td>
<td>Inside</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail</td>
<td>696</td>
<td>Inside</td>
<td>112</td>
<td>289</td>
<td>151</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>289</td>
<td></td>
<td>151</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1km</td>
<td>144</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3,105</td>
<td>Inside</td>
<td>491</td>
<td>1,279</td>
<td>627</td>
<td>708</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>1,279</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>627</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1km</td>
<td>708</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subgroups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Services</td>
<td>231</td>
<td>Inside</td>
<td>38</td>
<td>85</td>
<td>53</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>85</td>
<td></td>
<td>53</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>53</td>
<td></td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1km</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food ‘off premises’</td>
<td>119</td>
<td>Inside</td>
<td>17</td>
<td>43</td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>43</td>
<td></td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1km</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food ‘on premises’</td>
<td>127</td>
<td>Inside</td>
<td>17</td>
<td>63</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>63</td>
<td></td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1km</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food Retail</td>
<td>100</td>
<td>Inside</td>
<td>26</td>
<td>36</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400m</td>
<td>36</td>
<td></td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600m</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1km</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.3.2 Measuring the density of commercial addresses

This section discusses the method for calculating the density of commercial addresses within and around each of the 12 LSOA study areas. The first step was to map the boundaries of the study areas and the commercial addresses in Stoke-on-Trent within a GIS. A buffer analysis was used to create concentric boundaries every 200 metres up to a distance of 1km around the edge of each LSOA. Then a point in polygon (common GIS tool) analysis was used to record the number of addresses within each LSOA and within each 200m buffer. These counts give a general profile of the amount of commercial destinations within and around each area. These measures help to quantify the fact that the 12 LSOAs in the study are in very different types of areas in terms of the level of residential and commercial mix.

In addition to the raw counts of commercial addresses the density per square kilometre was calculated. This was important because the LSOAs differ in physical size. This was done by using the GIS software to calculate the land area within each LSOA and within each 200m buffer.

Table 7.2 demonstrates that the Stoke study area has a much higher density of retail addresses within the LSOA itself compared to the other study areas. However, looking at the counts at 500m shows that other areas, such as Hartshill and Penkhull, also have a higher than average density of retail addresses. This shows the potential importance of looking at relationships between commercial density and other measures for more than one neighbourhood buffer distance.

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Retail premises per sq.km within:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LSOA</td>
</tr>
<tr>
<td>Sneyd Green</td>
<td>6.28</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>4.73</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>18.96</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>28.99</td>
</tr>
<tr>
<td>Hanford</td>
<td>5.35</td>
</tr>
<tr>
<td>Stoke</td>
<td>131.13</td>
</tr>
<tr>
<td>Adderley</td>
<td>5.49</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>0.00</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>14.24</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>26.78</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>13.36</td>
</tr>
<tr>
<td>Trentham</td>
<td>25.37</td>
</tr>
</tbody>
</table>
7.3.3 Calculating the proximity of commercial addresses

The process of calculating proximity to commercial addresses for the residential population within each study area involved the use of network analysis methods as described in section 2.2.2. The approach used was to create an origin-destination matrix of distances between residential address locations (defined in Chapter 4) and some of the commercial address locations identified above in section 7.3.1. The distances calculated were measured along the road and pathway network built in Chapter 5. Distances were measured up to a maximum of 1km.

These proximity measures are produced to give an indication of walking distances to commercial destinations. As stated in the aims of this chapter, this focused on the types of commercial locations that are more likely to receive regular visits by neighbouring populations. For this purpose, only commercial premises classified above as local services, retail, eating and drinking and entertainment were included in this analysis.

The creation of this origin-destination matrix allowed the following measures to be calculated:

For each residential address:

- Distance to the nearest commercial premises (or food outlet) by type.
- Number of commercial premises (or food outlets) within x metres by type.

For each LSOA:

- Average distance from residential addresses to the nearest commercial premises (or food outlet) by type. (Also minimum, maximum and range of distances.)
- Population (and percentage of total) within 200m, 400m…1km of commercial premises (or food outlets) by type. As shown in Table 7.3.
- Population (and percentage of total) within x metres of 1, 2, 3… commercial premises (or food outlets) by type.
Table 7.3 Percentage of residential addresses within x metres of a fast food outlet

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Residential Addresses</th>
<th>% of addresses within</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>200m</td>
</tr>
<tr>
<td>Sneyd Green</td>
<td>686</td>
<td>0</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>595</td>
<td>14.5</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>640</td>
<td>21.9</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>873</td>
<td>24.4</td>
</tr>
<tr>
<td>Hanford</td>
<td>714</td>
<td>0</td>
</tr>
<tr>
<td>Stoke</td>
<td>900</td>
<td>46.9</td>
</tr>
<tr>
<td>Adderley Green</td>
<td>489</td>
<td>11.7</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>671</td>
<td>2.1</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>646</td>
<td>14.9</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>732</td>
<td>7.7</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>675</td>
<td>17.2</td>
</tr>
<tr>
<td>Trentham</td>
<td>562</td>
<td>0</td>
</tr>
</tbody>
</table>

7.4 Discussion

The key dataset used here to locate commercial addresses was AddressLayer 2. In terms of location this is a very good dataset. The other key use of this dataset was to classify address locations. For this purpose the data occasionally has inconsistencies between its different classifications and also because commercial properties are often subject to change, the data can become out of date. However, this is considered only a minor problem that will affect a very small proportion of the addresses. An address may change in terms of the company that is present but very often the type of use of the premises stays the same due to the property’s designated planning use. It is certainly not feasible to ground truth all of the address locations for a study of this size and overall this was not deemed to be necessary.

The series of metrics produced by this methodology are considered to be fit for purpose for this type of study. They successfully quantify the variability of proximity to commercial destinations throughout the study areas. The pattern of results matched those expected by the research team based on knowledge of the study areas. These metrics can be used to examine relationships with other environmental measures and health outcomes data. In addition these objective metrics can provide a direct comparison to the recorded results in a community survey; for example, with questions that ask respondents to report on their perceived neighbourhood, e.g. how far they believe it is to their nearest shop, food outlet etc.
Developing this analysis further would involve examining the choice and quality of food available in more detail and determining how far people are willing to travel for different types of food or different types of store.
8 Locating environmental supports for physical activity

8.1 Locating physical activity facilities

8.1.1 Aim

Facilities, places and public spaces that can be used to facilitate physical activity are a major resource for neighbourhoods in terms of increasing PA levels in the population. The locations of PA facilities need to be identified before the level of access can be measured in a GIS environment. In addition, the type of facility and its availability to the local population need to be recorded. Facilities with the ‘potential’ to be used for Physical Activity, such as community halls were also located. This section highlights data sources that can be used for this purpose and how they can be combined to produce a definitive list of facilities.

8.1.2 OS MasterMap AddressLayer 2

The AddressLayer 2 dataset was used to identify, locate and partially classify PA facility addresses for use within the study. This dataset provides the geographic point location (easting and northing) of each postal address within the UK to the nearest 1 metre.

Each address has been classified by Ordnance Survey into a Base classification which gives an indication of the type or use of an address. This Base classification has been grouped into two commonly used classifications from the National Land Use Database (NLUD) and the Valuation Office Agency (primary description field and special category codes).

More detailed information regarding Address Layer 2 can be found in Appendix 2.

The classifications include sites that have the potential to act as a PA facility, such as community / church halls. These categories were used to broadly classify PA facilities by type as:

- Physical Activity facilities (this includes purpose built facilities and activities located in community halls)
- Sports facilities (subset): these are facilities or halls that allow participation in sports.
- Activity-based facilities (subset): these are facilities or halls that allow participation in a particular activity, i.e. gym, dance.
- Potential facilities: these are community halls that have the potential to be used for physical activity.
8.1.3 Local Authority data and schools

Lists of Local Authority run facilities were obtained in order to act as a check with other data sets. A list of all swimming pools and sports facilities provided the main list of PA facilities and this helped to classify or reclassify the same facilities as found in OS MasterMap AddressLayer 2.

In addition, lists of community halls and youth centres were gathered and these could also be used to add to the list of potential PA facilities found in AddressLayer 2 and to classify addresses already found.

A complete list of schools within our study areas was also collected. These schools could then be found and mapped using AddressLayer 2. A number of schools in Stoke-on-Trent provide sports and other PA facilities to the public out of normal school times and therefore it was important to include them in any analysis.

All of the data collected from the local authority contained full or partial contact details for each facility which enabled checking of the type of activities available.

8.1.4 Combining data sources and classifying PA facilities

Data from the Local Authority and AddressLayer 2 was matched together and for every facility the correct classification was applied. If the classification recorded by the Local
Authority was different from AddressLayer two then the Local Authority classification was used.

The next step was to contact (by telephone) each facility to determine the type of physical activities available at a location and whether the facility is open to the general public and whether there was a charge / fee for using the facility.

These steps were necessary to produce a comprehensive list of PA facilities and to correctly classify them in terms of type of activity and public access.
8.2 Identifying areas of green space

8.2.1 Aims

Green space is important for helping to facilitate participation in physical activity behaviours within a neighbourhood. Green space primarily offers a potential site for a variety of physical activities. Green space also acts as a potential destination for walking trips. Areas of green space that have pleasant landscape and visual features are attractive areas to visit. In addition to this, increased levels of attractive green space in a neighbourhood can have an impact on the perceptions of the population and make the neighbourhood as a whole a more inviting place to walk or exercise.

8.2.2 Local Authority data

The GIS boundaries of green space areas within the city were obtained from the Local Authority. This data was produced for the city council as part of its green space audit, which all local authorities in England have to carry out.

In addition, the data contains four key classifications. Firstly, the general typology of the green space is reported, e.g. park, outdoor sports, semi-natural, agriculture. In addition, the primary and secondary purpose of the green space is reported, and includes: biodiversity, formal recreation, informal recreation, landscape and visual, operational and redundant. Finally the general access level of each area is reported, these are grouped as unrestricted, limited or restricted.

Each of these classifications can be used to disaggregate any analysis. The primary and secondary purposes, along with access, are particularly useful to understanding the type of benefit green space is bringing to the neighbourhood, e.g. visual impact or PA space.

In terms of green space being used as a site for PA it is important to examine the size of the space and its functionality. Areas that were above a particular size threshold were selected as a subset because they have sufficient space for group physical activity to take place. The level of this size threshold could be determined locally. This study has disaggregated the analysis to focus on areas of green space that are larger than 2 hectares. The size two hectares is a common standard (Box and Harrison, 1993. Cole and Bussey, 2000) and is used in the Accessible Natural Greenspace Standard (ANGSt) defined by Natural England. Figure 8.1 provides an example of the green space mapping and level of detail included.
8.2.3 Locating access points

Measuring access to an area of green space is more complex than measuring access to a point resource such as a PA facility. Some green space facilities are very large, e.g. parks, and often have fencing or natural barriers around their perimeter. Therefore it is insufficient to measure access from a population source to the nearest green space boundary.

To overcome this limitation the access points to each area of green space have been mapped. Sites with no access points have been excluded from analysis looking at accessibility. However, these sites may still be included in other measures, such as ‘total area of green space’ because an area that is inaccessible can still provide a benefit to a neighbourhood, e.g. areas fenced off for biodiversity protection could be bringing significant visual benefit.

Access points were mapped using on-screen digitising from aerial photography and from ground surveys. If an area of green space had open access along a particular boundary then an access point was digitised every 10 metres along the boundary. An example of mapped green space access points can be seen in Figure 8.2. This figure also shows pathways across green space areas.
8.3 Measuring access to environmental supports for physical activity

This section describes the methodologies used to produce metrics for the presence and accessibility of areas and facilities that support physical activity.

8.3.1 Density of PA facilities

The density of PA facilities around a population source was calculated using a point in polygon methodology to count the facilities within an area identified using a buffer. Buffer distances used were calculated from 200m to 1km at 200m intervals as well as a 500m buffer. The counts of PA facilities (broken down by type) were divided by the area of the buffers and the results were reported as facilities per square kilometre.

These buffers were applied to LSOA study areas, census output areas and to every residential address location as described in Chapter 4.

8.3.2 Density of green space

The density of green space was reported by calculating the total area of green space by type within set buffer distances up to 1km. Therefore, the percentage of the total area
Within each buffer that is green space was calculated and broken down by type and access.

### 8.3.3 Proximity to PA facilities and green space

Proximity to PA facilities and to green space along the pedestrian network was calculated using network analysis methodologies described in section 2.2.2. An origin-destination matrix was created to report the distance from every residential address location (located in Chapter 4) and every PA facility. A further origin-destination matrix was created to report the distance between every residential address location and every green space access point. In the case of green space this results in the distance to multiple access points of the same green space area being reported. In these cases only the closest access point was used for each area of green space. Figure 8.3 provides an example of address classified by proximity to ‘unrestricted access’ green space.

**Figure 8.3 An example of the green space accessibility analysis**

The green space proximity analysis highlighted the importance of disaggregating data by size and access. Table 8.1 shows the effect on results when looking at subsets of green space.
Table 8.1 An example of green space accessibility results

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Percentage of population within 300m of:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All green space</td>
<td>Green space greater than 2 hectares in size</td>
<td>Green space greater than 2 hectares in size and unrestricted</td>
<td></td>
</tr>
<tr>
<td>Sneyd Green</td>
<td>88.9</td>
<td>88.6</td>
<td>88.6</td>
<td></td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>97.8</td>
<td>79.4</td>
<td>79.3</td>
<td></td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>95.3</td>
<td>75.9</td>
<td>75.9</td>
<td></td>
</tr>
<tr>
<td>Heron Cross</td>
<td>75.9</td>
<td>46.4</td>
<td>46.4</td>
<td></td>
</tr>
<tr>
<td>Hanford</td>
<td>96.9</td>
<td>93.4</td>
<td>77.9</td>
<td></td>
</tr>
<tr>
<td>Stoke</td>
<td>97.8</td>
<td>28.2</td>
<td>20.1</td>
<td></td>
</tr>
<tr>
<td>Adderley Green</td>
<td>76.1</td>
<td>67.9</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Meir Hay</td>
<td>90.9</td>
<td>29.4</td>
<td>29.4</td>
<td></td>
</tr>
<tr>
<td>Trent Vale</td>
<td>79.9</td>
<td>32.0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>92.6</td>
<td>56.1</td>
<td>17.5</td>
<td></td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>60.1</td>
<td>50.4</td>
<td>49.6</td>
<td></td>
</tr>
<tr>
<td>Trentham</td>
<td>63.5</td>
<td>11.9</td>
<td>8.5</td>
<td></td>
</tr>
</tbody>
</table>

This example shows that when looking at all green spaces, there appears to be a high level of access in every study area. However, introducing a size of 2 hectares as a constraint dramatically reduces the number of people with access. For example, Meir Hay drops from 90.9% to 29.9% of the population within 300m. The same effect can be seen when adding a further constraint of ‘unrestricted access.’ For example the Trent Vale LSOA has dropped from 79.9% to 32% and finally 0% with access to ‘unrestricted’ green space greater than 2 hectares.

8.4 Discussion

The dataset used here to locate PA facilities once they had been identified from other sources was AddressLayer 2. The other key use of the datasets was to classify facilities. For this purpose the data sets occasionally have inconsistencies between them or a facility may have multiple functions. The only way to obtain a detailed picture of the type of facility and levels of public access was to contact each facility in turn. This is a very time consuming process.

In terms of green space the location data obtained from the local authority was very good. The classifications reported were useful in determining the general purpose of a green space but more work could be done on capturing levels of functionality and usability e.g. condition. Also, green space classifications used elsewhere will not necessarily be consist-
ent, which could make it difficult to compare results with other studies in terms of type of green space etc.

The series of metrics produced by this methodology are considered to be fit for purpose for this type of study. They successfully quantify the variability of proximity to PA facilities and green space throughout the study areas. In addition these objective metrics can provide a direct comparison to the recorded results in a community survey; for example, with questions that ask respondents to report how far they believe it is to their nearest leisure centre or park etc.
9 Calculating land use mix

9.1 Aims

This chapter discusses a methodology for producing a measure of land use mix within each study area. Evidence from the USA suggests that people who live in high density mixed land use areas are more likely to take part in physical activity and less likely to be obese (Ewing et al., 2003). The general concept is that neighbourhoods which enjoy a wide mix of land use are more likely to stimulate physical activity in the population. In other words, a variety of land use provides a number of different potential destinations for walking trips for recreation or everyday activities. An area containing a mix of residential, retail and green space land use provides more scope for physical activity than an area that is dominated by residential land use.

9.2 Data sources

In the United Kingdom there are very few good quality land use datasets, particularly data that covers the whole country. A number of local authorities have collected data as part of the National Land Use Database. However, this is a point location dataset and is not easy to use for this particular study, where the area of land use in different groups is required for a defined area.

9.2.1 Generalised Land Use Database Statistics for England

Generalised Land Use Database Statistics for England provide land use mix data for Lower Level Super Output Areas (LSOAs - the unit used to define the study areas) and Census Output Areas. These are experimental statistics representing 2005; they are designed to help increase the evidence base for neighbourhood renewal and sustainable communities (DCLG, 2007).

This dataset has been produced by using an automated method to classify OS MasterMap Topography Layer polygons. A detailed description of this methodology can be found in the DCLG Report 2007. Further details of OS MasterMap can be found in Appendix 2.

The dataset provides the area of land use for the categories listed in Table 9.1; some of these categories have been combined for the purpose of this study.
Table 9.1 Land use classification categories

<table>
<thead>
<tr>
<th>GLUD categories</th>
<th>Categories used in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic buildings</td>
<td>Residential</td>
</tr>
<tr>
<td>Domestic gardens</td>
<td></td>
</tr>
<tr>
<td>Non-domestic buildings</td>
<td>Commercial</td>
</tr>
<tr>
<td>Roads</td>
<td>Transport</td>
</tr>
<tr>
<td>Paths</td>
<td></td>
</tr>
<tr>
<td>Rail</td>
<td></td>
</tr>
<tr>
<td>Greenspace</td>
<td>Green space</td>
</tr>
<tr>
<td>Water</td>
<td>Other</td>
</tr>
<tr>
<td>Other land uses (largely hardstanding)</td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td></td>
</tr>
</tbody>
</table>

Due to the way this data has been produced there may be errors in the results. The data contains three figures which help to quantify this potential error. Firstly, the total land area for all categories is added together and reported. Secondly, the known area of each administrative area is reported. Finally, the percentage difference between these two reported areas is calculated. This figure gives an indication of whether land use is being under reported or over reported in each area.

9.3 Methods for calculating land use mix

9.3.1 Herfindahl-Hirschman Index

The Herfindahl-Hirschman Index (HHI) (AmosWEB, 2004) is a measure developed originally in the field of economics to measure the concentration of market shares between firms or sectors. The HHI is simply the sum of squares of the percentages of each land use type in an area; the index can range between 2,000 (given that we have 5 categories of land use) and 10,000. For example, if there is only one land use type in an area the resulting index will be 10,000. Therefore the greater the index the lower the level of land use mix.

The Generalised Land Use Database figures were used to calculate the percentage of land use in each category. These figures were then used to calculate the HHI. Table 9.2 shows the resulting figures for each LSOA within the study. These figures were also calculated for each Census Output Area. In areas demonstrating a low level of mix it was also important to identify the dominant form of land use.
Table 9.2 HHI results for each study area

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Percentage of land use by category</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>HHI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Residential</td>
<td>Commercial</td>
<td>Transport</td>
<td>Green space</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sneyd Green</td>
<td>50.69</td>
<td>1.09</td>
<td>13.32</td>
<td>31.82</td>
<td>3.07</td>
<td>3,771</td>
<td></td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>49.04</td>
<td>2.71</td>
<td>15.18</td>
<td>26.58</td>
<td>6.50</td>
<td>3,391</td>
<td></td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>46.58</td>
<td>2.38</td>
<td>20.45</td>
<td>25.28</td>
<td>5.31</td>
<td>3,261</td>
<td></td>
</tr>
<tr>
<td>Heron Cross</td>
<td>12.70</td>
<td>11.31</td>
<td>16.11</td>
<td>38.62</td>
<td>21.26</td>
<td>2,492</td>
<td></td>
</tr>
<tr>
<td>Hanford</td>
<td>29.86</td>
<td>0.31</td>
<td>8.87</td>
<td>58.33</td>
<td>2.63</td>
<td>4,380</td>
<td></td>
</tr>
<tr>
<td>Stoke</td>
<td>20.79</td>
<td>14.21</td>
<td>20.28</td>
<td>26.56</td>
<td>18.17</td>
<td>2,081</td>
<td></td>
</tr>
<tr>
<td>Adderley Green</td>
<td>34.37</td>
<td>2.75</td>
<td>9.82</td>
<td>31.55</td>
<td>21.50</td>
<td>2,744</td>
<td></td>
</tr>
<tr>
<td>Meir Hay</td>
<td>74.50</td>
<td>0.07</td>
<td>21.40</td>
<td>3.59</td>
<td>0.45</td>
<td>6,021</td>
<td></td>
</tr>
<tr>
<td>Trent Vale</td>
<td>54.46</td>
<td>0.92</td>
<td>18.90</td>
<td>22.19</td>
<td>3.53</td>
<td>3,829</td>
<td></td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>54.96</td>
<td>5.74</td>
<td>18.18</td>
<td>9.38</td>
<td>11.74</td>
<td>3,610</td>
<td></td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>37.05</td>
<td>0.98</td>
<td>12.02</td>
<td>39.35</td>
<td>10.59</td>
<td>3,179</td>
<td></td>
</tr>
<tr>
<td>Trentham</td>
<td>71.56</td>
<td>1.10</td>
<td>16.97</td>
<td>8.10</td>
<td>2.27</td>
<td>5,480</td>
<td></td>
</tr>
</tbody>
</table>

The Stoke LSOA has a very good mix of land use. The highest percentage of any land use category is 26.56% (green space) and therefore the HHI score is very close to 2,000. In the Meir Hay LSOA the dominant land use is residential (74.5%) followed by transport (21.4%) resulting in a HHI score of 6,021.

9.3.2 Spatial entropy

A number of studies outside of the UK have used a spatial entropy equation to generate an entropy score for each area of a study (Frank et al 2005).

The score is calculated using the following formula, where \( k \) is the category of land use; \( p \) is the proportion of land area devoted to a specific land use; \( N \) is the number of land use categories:

\[
\frac{- \sum_{k} (p_k \ln p_k)}{\ln N}
\]

This produces a score ranging from 0 to 1, with 0 representing homogeneity and 1 representing heterogeneity. The results for each LSOA study area are reported in Table 9.3 and are reported next to the HHI results.
Table 9.3 Spatial Entropy results for each study area

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Spatial Entropy Score</th>
<th>HHI Score</th>
<th>Spatial Entropy Rank</th>
<th>HHI Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sneyd Green</td>
<td>0.70</td>
<td>3,771</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>0.78</td>
<td>3,391</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>0.79</td>
<td>3,261</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>0.93</td>
<td>2,492</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Hanford</td>
<td>0.62</td>
<td>4,380</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Stoke</td>
<td>0.99</td>
<td>2,081</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Adderley Green</td>
<td>0.86</td>
<td>2,744</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>0.43</td>
<td>6,021</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>0.71</td>
<td>3,829</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>0.79</td>
<td>3,610</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>0.79</td>
<td>3,179</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Trentham</td>
<td>0.55</td>
<td>5,480</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

The Stoke LSOA has a score of 0.99, representing the fact that it has a very even mix of land use types. The Meir Hey LSOA is still reported as being the worst area in terms of level of land use mix with a score of 0.43. The rankings of the spatial entropy scores closely match those of the HHI scores. In the middle of the rankings there are some LSOAs that differ slightly in rank which is down to the effect of the underlying calculations.

9.4 Discussion

Both methods explored in this study produce an easy to understand metric of the level of land use mix. The results appear to capture the variability in land use seen on the ground.

The quality of the data held in the Generalised Land Use Database is variable and the methodology used to create it is quite simplistic. However, for this study, a crude measure of land use by broad categories is sufficient and therefore the data still provides a useful measure. Also, the data has been put together using the same process for the whole of England and therefore it helps to provide a measure of land use mix that can be directly compared to other studies in other parts of the country.

A limitation of using the area-based land use figures reported in the GLUD is that neighbourhood areas (buffer or service area boundaries) around a study site or address location may cut through a LSOA or census output area and therefore exact land use figures are
difficult to determine. A solution to this would be to use OS MasterMap Topographic layer polygons as land use parcels (if available) and to classify them using the same method described in the GLUD report (DCLG, 2007).
10 Mass Transport Provision

10.1 Aims

The chapter focuses on the methodology for measuring mass transport (public transport) provision for the residential population in a study area. For those who cannot afford a car, or are too young to drive, availability and proximity of mass transit may facilitate physical activity i.e. walking and “active transport”. This was approximated by calculating the distance to bus stops or train stations along the road and pathway network.

10.2 Data sources

This method uses the residential address locations identified in Chapter 4 and the roads and pathways identified in Chapter 5.

Detailed bus stop locations were provided in a GIS layer by Stoke-on-Trent City Council. These locations were surveyed by the local authority and provide very accurate locations that even show the side of the street that each bus stop is on. This is an important detail in areas where the ‘other’ side of the road is not easily accessible.

Stoke-on-Trent only has one major train station and only two minor stations going to a very limited number of destinations. Only one of these stations is within close proximity to any of the LSOA study areas. Therefore, for the purpose of this study the proximity to train stations has not been measured.

10.3 Methodology

The approach used was to use a network analysis approach (as described in section 2.2.2) to calculate the distance from every residential address location to the nearest bus stop. This produces a very simple measure of access to mass transport.

The figures for each residential address were used to generate the following metrics for each study area (as well as Census Output Areas):

- Average distance from residential addresses to the nearest bus stop (also minimum, maximum and range of distances).
- Population (and percentage of total) within 200m, 400m (up to a 1 km) of a bus stop. The results produced for this study can be seen in Table 10.1.
Table 10.1 Access to a bus stop along the road and pathway network

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Percentage of population within 200m of a bus stop</th>
<th>Percentage of population within 400m of a bus stop</th>
<th>Average distance (m) from each address to the nearest bus stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sneyd Green</td>
<td>34</td>
<td>98</td>
<td>233</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>53</td>
<td>100</td>
<td>178</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>70</td>
<td>95</td>
<td>168</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>73</td>
<td>98</td>
<td>151</td>
</tr>
<tr>
<td>Hanford</td>
<td>62</td>
<td>100</td>
<td>167</td>
</tr>
<tr>
<td>Stoke</td>
<td>83</td>
<td>100</td>
<td>134</td>
</tr>
<tr>
<td>Adderley Green</td>
<td>77</td>
<td>100</td>
<td>140</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>30</td>
<td>85</td>
<td>276</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>78</td>
<td>100</td>
<td>143</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>61</td>
<td>100</td>
<td>167</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>53</td>
<td>99</td>
<td>198</td>
</tr>
<tr>
<td>Trentham</td>
<td>40</td>
<td>72</td>
<td>298</td>
</tr>
</tbody>
</table>

As an alternative to the network analysis figures the density of bus stops in an area was also measured. This was reported by calculating the number of bus stops per km². These were calculated for each study LSOA (and OA) and each buffer area. An example of these measures is shown in Table 10.2 along with a comparison to the network distance results.

Table 10.2 Bus stop density and average distance to a bus stop

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Density of bus stops (per km²) within a 200m buffer of the study area</th>
<th>Average distance (m) from each address to the nearest bus stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sneyd Green</td>
<td>12</td>
<td>233</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>24</td>
<td>178</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>23</td>
<td>168</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>14</td>
<td>151</td>
</tr>
<tr>
<td>Hanford</td>
<td>13</td>
<td>167</td>
</tr>
<tr>
<td>Stoke</td>
<td>35</td>
<td>134</td>
</tr>
<tr>
<td>Adderley Green</td>
<td>17</td>
<td>140</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>26</td>
<td>276</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>19</td>
<td>143</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>31</td>
<td>167</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>15</td>
<td>198</td>
</tr>
<tr>
<td>Trentham</td>
<td>14</td>
<td>298</td>
</tr>
</tbody>
</table>
A further alternative to this measure is to calculate the number of bus stops per 1,000 of population in an area. This type of measure and the density method provide an alternative approach to approximating mass transport provision which would be suitable for a study that did not have detailed road and pathway data.

10.4 Discussion

The key metric produced here to measure the distance to the nearest bus stop is a very simplistic approach. However, it does illustrate the differences in access, both between (Table 10.1) and within areas and it is considered appropriate for this type of study.

The reason it is so simplistic is because it only looks at the closest stop. In terms of participation in physical activity, the proximity of a bus stop provides a ‘walking’ destination. However mass transport can play a part in getting people to a destination where they can be physically active or where there is a need to continue on foot to get to their final destination. Furthermore levels of car ownership are low in Stoke-on-Trent (70%) compared to the national average and 8 percent of all travel movements are by bus in the city (City Council, Local Transport Plan, 2006).

Local Authorities often use such an approach to calculate levels of access, using public transport, to key facilities such as hospitals. If detailed information on bus routes is known then this approach provides a potentially interesting way of looking at access to physical activity opportunities such as parks and open space.
11 Road Traffic and Street Safety

11.1 Aims

This chapter aims to demonstrate a method for producing metrics of road traffic and street safety. These are important measures because they can both have a significant impact on individuals’ perceptions of their neighbourhood environment and can act as a potential barrier to participation in physical activity.

11.2 Road Traffic Levels

11.2.1 Data sources

The most suitable road traffic data available for Stoke-on-Trent was data collected by the Local Authority. The data is collected in the following two ways:

- Automatic traffic count stations - An automatic traffic count is carried out at each location for a period of 7 days. This data is then analysed by the city council and a 5-day (Monday to Friday) and 7-day (Monday to Sunday) daily average of traffic count flows is produced.

- Detailed manual counting - Manual counting is only carried out for a period of 1 day.

In both cases data was obtained for the last 5 years (January 2002 to December 2006). During this time frame, some locations were surveyed on an annual basis (but not necessarily at the same time every year) and others were surveyed only once. The locations that were surveyed every year tend to be the routes with the highest traffic flows. In Stoke-on-Trent, during the last 3 years, there have been two areas that have seen major long term road works. This type of long term work can have a significant impact on the traffic levels of neighbouring streets. However, following these works traffic levels in neighbouring areas would be expected to return to normal lower levels. For this reason, it is important for any study looking at localised traffic counts to examine data over a number of years in order to identify areas with a temporary significant increase in levels.

The location of the traffic counts was recorded by the surveyors as an easting and northing reported to the nearest 100 metres. In addition to this the data reports the name of the road.
11.2.2 Methodology

The approach used to produce a useful metric of traffic levels for each study area is a labo-
rious manual process. An approach of calculating the average traffic levels from all sample
points in an area would not produce a useful figure. The method used had the following
stages:

i) The first stage of the methodology involved mapping the point locations of traffic
counts using the geographic coordinates provided by the city council. This gives
an initial picture of the spatial distribution of the sample points.

ii) The second stage of the process was to resolve which traffic counts would be used
at each location. From the last five years of data, the most recent figure available
at each location was chosen. However at locations where there is known to be
a high recent count due to temporary activities then earlier figures can be used.
If both automatic and manual counts were available at the same location, the
automatic counts were used because they were recorded over a longer time frame
and should be more reliable.

iii) The third and most problematic stage is to assign traffic counts to a digital road
network. The road network was mapped using the OS MasterMap Integrated
Transport Network layer as described in Chapter 5. The next step was to assign
to each section of road the traffic counts of the nearest sample point, but only if
the name of the road (or road number) matched for both sets of data. A number
of streets do not have sample points on or near to them. In the majority of
cases these streets are in residential areas and are assumed to have very low
traffic made up largely from residents of the street. In a few cases there were
roads that had no information but they were part of non residential areas. These
roads remained unclassified and are assumed to have traffic levels associated with
the local commercial activity.

Street parking
The fourth stage requires the traffic levels for each road to be grouped into categories. This was done by ranking the chosen counts for all survey points in the study and grouping the values by quintile. This allowed the classification of roads as shown in Table 11.1. In addition to assigning quintiles to the roads, a subset of roads with greater than 500 vehicles per day was measured. These roads represented arterial routes with greater than local levels of traffic. Figure 11.1 provides a mapped example of this for the Stoke study area.

Table 11.1 Road traffic quintile classifications

<table>
<thead>
<tr>
<th>Quintile</th>
<th>Average vehicles per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13,430 plus</td>
</tr>
<tr>
<td>2</td>
<td>4,587 to 13,429</td>
</tr>
<tr>
<td>3</td>
<td>2,015 to 4,586</td>
</tr>
<tr>
<td>4</td>
<td>545 to 2,014</td>
</tr>
<tr>
<td>5</td>
<td>544 or less</td>
</tr>
</tbody>
</table>

Figure 11.1 An example of mapped road traffic classifications
v) The final stage of the methodology was to calculate the length of road in each category for each study area and for buffered areas around each study area. The results of this were then standardised to record the percentage of total road length in each traffic quintile. Table 11.2 gives an example of this for the LSOA study areas.

Table 11.2 Road Traffic indices for LSOA study areas

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Percentage of total road length in each traffic quintile</th>
<th>Percentage of roads with no recorded traffic (assumed low)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highest 1</td>
<td>2</td>
</tr>
<tr>
<td>Sneyd Green</td>
<td>5.58</td>
<td>6.22</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>9.30</td>
<td>7.04</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>6.22</td>
<td>5.53</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>14.48</td>
<td>5.34</td>
</tr>
<tr>
<td>Hanford</td>
<td>17.16</td>
<td>8.13</td>
</tr>
<tr>
<td>Stoke</td>
<td>13.87</td>
<td>7.23</td>
</tr>
<tr>
<td>Adderley Green</td>
<td>9.76</td>
<td>3.07</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>4.42</td>
<td>9.37</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>5.14</td>
<td>13.99</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>11.88</td>
<td>5.73</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>9.21</td>
<td>5.91</td>
</tr>
<tr>
<td>Trentham</td>
<td>7.50</td>
<td>9.30</td>
</tr>
</tbody>
</table>

11.2.3 Discussion

The measures produced by this method provide a good proxy of traffic levels within the neighbourhood of each study area. They provide useful figures to compare to the population’s perceptions of traffic with their neighbourhood.

However, the issue of road traffic as a barrier to physical activity is a complicated one. It may be true in some areas that traffic speeds act as more of a barrier to physical activity. This could be particularly true in some suburban and rural areas, where the general flow of traffic is relatively low but the speed at which traffic moves through an area is relatively high and therefore affects certain group’s perceptions of the safety of an area.

For this particular study, good quality data reporting traffic speeds was not available. However, a general approach would be to use the type of road and associated speed limit as a proxy. This study is in an urban setting and therefore the speed of traffic is believed to
be less of an issue. These measures do not account for the amount of parked traffic in an area, which can act as an obstacle and safety concern for pedestrians and reduce areas for safe play.

11.3 Accident Data

11.3.1 Data sources

A dataset providing the geographically referenced point location of road accidents was provided by Stoke-on-Trent City Council. The data supplied represented the last three full years of available data (2003 to 2005).

This is a comprehensive dataset and allows analysis of accidents in an area to be disaggregated in a number of ways. In addition to the location of the accident, each point provided the following detail:

- The date, day of week and time of day of the accident.
- The number of people involved.
- The gender and age of each person involved.
- The severity of injury to each casualty (fatal, serious, minor)
- The mode of transport used by each casualty (car, motorcycle, cyclist, pedestrian and others)

11.3.2 Methodology

The method first involved mapping the point locations of each accident in Stoke-on-Trent, particularly the accidents within 1 kilometre of each study area.

A simple buffer and point in polygon analysis was then used to identify the number of accidents (and casualties) within each study area and within set distances of each study area up to a distance of 1km. Figure 11.2 shows accident location mapped along with traffic data for the Stoke LSOA.
These total counts were then disaggregated by type and severity. If required the results can be disaggregated by the other categories held in the data and listed above.

In addition to these raw counts the data was standardised in two ways. Firstly, the counts per square kilometre were calculated using the recorded area of each study area and buffered area. Secondly, the number of accidents per kilometre of road was calculated by using the recorded total length of road in each study area and buffer area. An example of these two measures can be seen in Table 11.3.
Table 11.3 Accident casualty measures

<table>
<thead>
<tr>
<th>Study site</th>
<th>Casualties per sq.km</th>
<th>Casualties per km of road</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All casualties</td>
<td>Pedestrians or cyclists</td>
</tr>
<tr>
<td>Sneyd Green</td>
<td>8.60</td>
<td>2.15</td>
</tr>
<tr>
<td>Sandford Hill - West</td>
<td>19.71</td>
<td>4.13</td>
</tr>
<tr>
<td>Sandford Hill - East</td>
<td>22.15</td>
<td>3.76</td>
</tr>
<tr>
<td>Heron Cross</td>
<td>22.93</td>
<td>3.91</td>
</tr>
<tr>
<td>Hanford</td>
<td>6.25</td>
<td>0</td>
</tr>
<tr>
<td>Stoke</td>
<td>26.05</td>
<td>4.70</td>
</tr>
<tr>
<td>Adderley Green</td>
<td>7.24</td>
<td>1.45</td>
</tr>
<tr>
<td>Meir Hay</td>
<td>16.75</td>
<td>3.09</td>
</tr>
<tr>
<td>Trent Vale</td>
<td>11.29</td>
<td>2.26</td>
</tr>
<tr>
<td>Hartshill and Penkhull</td>
<td>14.84</td>
<td>2.34</td>
</tr>
<tr>
<td>Baddeley Green and Milton</td>
<td>6.12</td>
<td>1.70</td>
</tr>
<tr>
<td>Trentham</td>
<td>4.79</td>
<td>1.31</td>
</tr>
</tbody>
</table>

11.3.3 Discussion

The metrics produced using this methodology provide indicators of the number and type of accidents. A further approach would be to identify hot spots or localities where there are a significantly high number of accidents and then examine the proximity of the population to these hotspots. Similarly, accident data could be used to score streets as low, medium and high probability of accidents and by type of accidents.

However, a number of problems exist when using street accident data. Firstly, there are often relatively low counts of accidents in areas, particularly when broken down into categories of type and severity.

Secondly, road accidents are being measured as a potential barrier to participation in physical activity (walking and cycling) along a route because they may deter people from using a route by affecting their perceptions of safety. With this in mind it is difficult to quantify the effect of accidents in an area on perceptions of safety. This is because it is difficult to determine which has the greater impact, is it frequency of accidents or is it the severity? The temporal element of impact on perceptions is also important.
12 Crime Statistics

12.1 Aims

This chapter demonstrates a method for producing metrics of crime and anti-social behaviour. Both crime and anti-social behaviour are potential barriers to participation in physical activity. Individuals’ perceptions of safety in their neighbourhood will have an impact on whether they decide to engage in everyday activities such as walking to a shop or walking through a park. This may be particularly true for vulnerable groups within the population.

12.2 Data sources

Data is freely available through the Office of National Statistics at an area level (including LSOAs) for broad categories of crime throughout the country. However, in order to be able to measure crime in and around the study LSOA areas the actual point locations of crimes were required; this data was acquired from Staffordshire Police and the City Council using a formal data sharing agreement. This process was necessary because of the sensitive nature of the data. A number of restrictions were placed on the use of the data, such as, only gaining access to necessary attributes (no data on individuals) and not publishing maps or other information that show exact locations of crimes. These restrictions did not prevent the measures required in this study from being created because data is aggregated for each study area.

The types of crime recorded in the data vary from a violent incident to a case of fraud. For the purposes of this study only crime believed to have a potential effect on neighbourhood perceptions of safety was requested. These incidents fell into 3 broad categories:

- Criminal Damage
- Violent Crime
- Burglary of Dwellings

The data used provides a geo-referenced point location for every reported crime in the last 3 years within Stoke-on-Trent. The attributes of these crimes included a general grouping of the type of crime, a specific description code of the type of crime and the date on which the crime was reported.

In addition, a geo-referenced list of point locations for anti-social behaviour incidents was provided for the same time period. This data included a type description of each incident and the date reported.
12.3 Methodology

The metrics produced for this study were calculated for crime data recorded in 2006 and for crimes recorded in the 12 month period before the study (post July 2006).

The first step involved mapping the point locations of each incident of crime and anti-social behaviour in Stoke-on-Trent. The incidents occurring within 1km or each study area were then selected for analysis.

A simple buffer and point in polygon analysis was then used to identify the number of incidents within each study area and within set distances of each study area up to a distance of 1km. This was also done for each census output area.

These total counts of incidents of crime were then disaggregated by the 3 broad categories listed above (section 12.2). Anti-social behaviour counts were not broken down by category because only a general prevalence of incidents was required for this study.

The final step was to standardise to counts of incidents by area to produce the number of incidents per sq. km (density). These densities were calculated for each study LSOA (and OA) and for each buffer area around the study areas.

12.4 Discussion

The metrics produced using this methodology produce indicators of the number and type of crime and anti-social behaviour incidents. The results for the study areas did highlight differences between neighbourhoods and in general, areas with a greater level of land use mix, particularly commercial land use, appeared to have higher levels of both crime and anti-social behaviour.

As with other barriers to physical activity, a further approach would be to identify hot spots or localities where there are a significantly high number of incidents and then look at the proximity of the population to these hotspots. In terms of localities, the level of crime and anti-social behaviour could be measured around green space areas or near to shops and services. Similarly, metrics could be used to score streets and pathways as low, medium and high levels of crime or anti-social behaviour.

Like traffic levels and accidents, crime and anti-social behaviour are being measured as a potential barrier to participation in physical activity (walking and cycling). They may deter people from using a particular route or locality by affecting their perceptions of safety. The effect of crime and anti-social behaviour on perceptions will vary between different groups in the population. The type and severity of a crime may also have a big impact on perceptions.
Metrics produced here provide a useful comparison to responses from the parallel community survey. These comparisons may help to better understand the effects of crime on perceptions of safety and ability to participate in physical activity within a neighbourhood.
References


Appendix 1 Index of Deprivation Domains

Income Deprivation Domain
The purpose of this Domain is to capture the proportion of the population experiencing income deprivation in an area.

- Adults and children in Income Based Job Seekers Allowance households (2001).
- Adults and children in Working Families Tax Credit households whose equivalised income (excluding housing benefits) is below 60% of median before housing costs (2001).
- Adults and children in Disabled Person’s Tax Credit households whose equivalised income (excluding housing benefits) is below 60% of median before housing costs (2001).
- National Asylum Support Service supported asylum seekers in England in receipt of subsistence only and accommodation support (2002).
- In addition, an Income Deprivation Affecting Children Index and an Income Deprivation Affecting Older People Index were created.

Employment Deprivation Domain
This domain measures employment deprivation conceptualised as involuntary exclusion of the working age population from the world of work.

- Unemployment claimant count (JUVOS) of women aged 18-59 and men aged 18-64 averaged over 4 quarters (2001).
- Participants in New Deal for the 18-24s who are not included in the claimant count (2001).
- Participants in New Deal for 25+ who are not included in the claimant count (2001).
- Participants in New Deal for Lone Parents aged 18 and over (2001).
Health Deprivation and Disability Domain
This domain identifies areas with relatively high rates of people who die prematurely or whose quality of life is impaired by poor health or who are disabled, across the whole population.

- Adults under 60 suffering from mood or anxiety disorders (1997-2002).

Education, Skills and Training Deprivation Domain
This Domain captures the extent of deprivation in terms of education, skills and training in a local area. The indicators fall into two sub domains: one relating to education deprivation for children/young people in the area and one relating to lack of skills and qualifications among the working age adult population.

- Sub Domain: Children/young people
  - Average points score of children at Key Stage 2 (2002).
  - Average points score of children at Key Stage 3 (2002).
  - Average points score of children at Key Stage 4 (2002).
  - Proportion of young people not staying on in school or school level education above 16 (2001).

- Sub Domain: Skills
  Proportions of working age adults (aged 25-54) in the area with no or low qualifications (2001).

Barriers to Housing and Services Domain
The purpose of this Domain is to measure barriers to housing and key local services. The indicators fall into two sub-domains: ‘geographical barriers’ and ‘wider barriers’ which also includes issues relating to access to housing, such as affordability.

- Sub Domain: Wider Barriers
  - LA level percentage of households for whom a decision on their application for assistance under the homeless provisions of housing legislation has been made, assigned to SOAs (2002).
Sub Domain: Geographical Barriers
- Road distance to GP premises (2003).
- Road distance to a supermarket or convenience store (2002).
- Road distance to a primary school (2001-2002).
- Road distance to a Post Office (2003).

Crime Domain
This Domain measures the incidence of recorded crime for four major crime themes, representing the occurrence of personal and material victimisation at a small area level.
- Burglary (4 recorded crime offence types, April 2002-March 2003).
- Theft (5 recorded crime offence types, April 2002-March 2003, constrained to CDRP level).
- Criminal damage (10 recorded crime offence types, April 2002-March 2003).
- Violence (14 recorded crime offence types, April 2002-March 2003).

The Living Environment Deprivation Domain
This Domain focuses on deprivation with respect to the characteristics of the living environment. It comprises two sub-domains: the ‘indoors’ living environment which measures the quality of housing and the ‘outdoors’ living environment which contains two measures about air quality and road traffic accidents.
Sub-Domain: The ‘indoors’ living environment

Sub-Domain: The ‘outdoors’ living environment
- Road traffic accidents involving injury to pedestrians and cyclists (2000-2002).

The methodological steps that were taken to create the IMD 2004 are described in the full report, available at http://www.communities.gov.uk/indices
Appendix 2 – OS MasterMap

AddressLayer 2
AddressLayer 2 is Ordnance Survey’s most detailed address data set. There are alternative data sets such as AddressPoint and AddressLayer that also contain detailed information on every address in Great Britain. All of these products could be used to successfully identify residential address locations.

AddressLayer2 contains additional Multi-occupancy without postal address (MOWPA) features which allows multiple households at a single address (such as flats) to be identified and thus providing a better representation of residential population (and population density).

Objects without a postal address (OWPA) provide the locations of a number of buildings such as community halls, and some features in the non-built environment such as playing fields and recreation grounds.

Integrated Transport Network (ITN) Layer
The ITN is a digital mapped road network representing all road centrelines to 1 metre accuracy, making it suitable for applications such as measuring the true length of a road and thus producing accurate accessibility measures.

The descriptive term field contains a classification of roads including motorways, A-roads, B-Roads, minor roads and local streets. The data also indicates whether the road is a slip road, a roundabout and is single or dual carriageway. Single carriageway roads are represented as a single line and dual carriageway roads as 2 lines, one for each side of the road.

The ITN contains a large amount of road routing information. These relate to traffic restrictions such as a no entry or one way street.

Topography Layer
The Topography Layer was not used in this study due to lack of availability and cost constraints. This layer provides a detailed spatial extent of features on the landscape including buildings, land areas, roads and water. This enables better visualisation of the urban environment. It also allows analysis of types of land use to be achieved with a high level of detail.

Further information about OS MasterMap including detailed technical documents can be found at: http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/