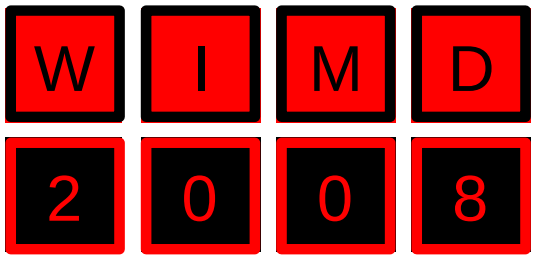
ystadegau ar gyfer cymru statistics for wales

Llywodraeth Cynulliad Cymru
Welsh Assembly Government



Welsh Index of Multiple Deprivation 2008

Technical Report

Contents

1 Introduction

This technical report describes how the Welsh Index of Multiple Deprivation (WIMD) 2008 was constructed. The deprivation domains which were used are listed, together with the indicators within each domain. The geographical units used are specified. The report also explains the statistical algorithms used to construct the index from these components.

The Welsh Index of Multiple Deprivation 2008 is the official measure of deprivation for small areas in Wales. It was developed for the Welsh Assembly Government by the Assembly's Statistical Directorate and the Local Government Data Unit (Wales). It replaces the index which was produced in 2005.

deprivation and poverty

Poverty and deprivation are regularly used interchangeably, but a clear distinction should be made between them. The condition of poverty means not having enough money or other financial assets or material possessions to meet needs. Deprivation refers to unmet need which is caused by a lack of resources in general, not just financial or tangible resources.

People can be considered to be deprived if they lack the material standards of diet and clothing, and do not participate in employment or recreation. If they lack the financial resources to obtain these conditions of life they are in poverty. It has been stated that while people experiencing some forms of deprivation may not all have low income, people experiencing multiple or single but very severe forms of deprivation are in almost every instance likely to have very little income and few or no other resources. So deprivation depends on the level of conditions and activities experienced, and poverty on the incomes and other material resources available to the individual or household.

A distinction has also been made between material and social deprivation: people may not have material goods, facilities, or amenities. Additionally they may not have access to ordinary social relationships and activities. If there are these different forms of deprivation, then some people may experience several forms of deprivation and others only a single form.

multiple deprivation

So it is clear that deprivation is a wider concept than poverty. As a summary, poverty means not having enough money (or other essentials) to get by. Deprivation refers to problems caused by a general lack of resources and opportunities (not just money).

Multiple deprivation has sometimes been thought of as a separate form of deprivation; one which could, in theory, be directly measured. It only required the discovery of a suitable metric, followed by a straightforward aggregation of a set of deprivation components. That is, calculating deprivation in the same way that a person's wealth might be worked out by adding up all their separate assets.

It is now generally accepted that multiple deprivation is a complex concept and not some separate form of deprivation. It is simply a combination of more specific forms of deprivation, which themselves can be measured to some extent. A key issue is whether combinations of these different forms of deprivation interact synergistically; that is, that they are not simply additive but have more impact if found in certain combinations.

If multiple deprivation cannot be directly measured then there are problems in technically validating any overall index, because technical validation requires something against which the index can be calibrated. It follows that the question of how components in the overall index might be weighted and combined becomes a pivotal issue.

levels of deprivation

Combining different deprivation attributes into a single index means that a number of issues relating to the links between differing forms of deprivation at the individual, household, and geographical area level have to be considered.

A significant issue is the extent to which individuals and families experiencing deprivation cluster together geographically, and the extent to which other individuals and families who could not be considered to be experiencing deprivation are affected by the overall level of deprivation in their area.

Although much of the data may be collected at the individual or household levels of deprivation, the results in any index are likely to be presented in the form of an aggregate score for some substantial geographical area. But this will combine deprivations experienced by many different groups within that area. With multiple deprivation it is not necessarily the same people experiencing the various forms of specific deprivation in an area. Some people will experience one kind of deprivation; others a different kind of deprivation. Some people will experience several kinds of deprivations; others none at all.

Finally there are several forms of deprivation that may be difficult to attribute on a precise geographical basis. For example, some aspects of crime (like vandalism or car crime in a city centre) might legitimately be associated with the areas of victims and potential victims who lived elsewhere, because they are potentially at risk if they work or shop in that city centre.

the approach to a solution

It is clear from any reputable study on deprivation that the financial component in any overall scale of deprivation is of great significance. This is taken into account in the weighting of the components which make up the overall index for the Welsh Index of Multiple Deprivation (WIMD) 2008. This was also true for the WIMD 2005, and for the similar indexes of the other UK countries.

It is also clear that the idea of separate forms of deprivation are differentially experienced by different groups. For this reason the idea of separate domains of deprivation has been developed.

The methodology adopted for the WIMD 2000, WIMD 2005 and WIMD 2008 focuses on the creation of a series of robust domain sub-indexes; that is, for income, health, education, and other factors separately. These domain sub-indexes can be validated, either because they directly measure the factor itself (for example, in the case of means-tested benefit, reliance on unemployment benefit), or by making comparisons with other research studies.

This approach results in a more complex index based on more data sources than earlier indexes of deprivation. It also requires domain-specific indicators; that is, indicators which are cogent measures of that domain's deprivations and are not just vaguely and imprecisely related to the domain. It also requires procedures for combining indicators and the related data within any domain according to well-defined algorithms. These procedures are described in Appendix C of this report.

The WIMD 2008 is made up of eight separate domains (or kinds) of deprivation:

- income
- employment
- health
- education
- housing
- environment
- access to services
- community safety

Each of them was based on a range of different indicators which meant that they were measured in different ways using different units. So before they could be combined the measurements had to be transformed to make them compatible (see Appendix D). For example, if the height of something had been measured in metres and the weight in kilograms it would not make sense simply to add them together.

Income and employment were classed as the most important factors, and they were given the biggest weighting in the overall index. Although there is more to deprivation than poverty, not having enough money or a job is a big part of it.

The weights used for WIMD 2008 were based on those for WIMD 2005, with the weights for the income, employment, health and education domains being reduced to create the new Community Safety domain.

Income	23.5%
Employment	23.5%
Health	14%
Education, skills and training	14%
Housing	5%
Physical environment	5%
Geographical access to services	10%
Community safety	5%

Development of the Index

Over the past forty years the development of indexes to measure deprivation at the local level has been regarded as an important feature of government policy, mainly for the identification of priority areas to focus regeneration and financial aid programmes more effectively. Originally these indexes were designed for specific policy areas; for example, health or education. But over the past two decades the concept of a general-purpose index of multiple deprivation at the local level has

become increasingly significant. The availability of electronic small area data (particularly the Census) contributed to this development.

The Welsh Index of Multiple Deprivation was designed to model levels of deprivation in Wales and support the objective targeting of resources. It used a more sophisticated methodology than the earlier Index of Socio-Economic Conditions and was based on more representative measures of deprivation at the small-area level, some of which used administrative data from new sources.

The 2000 index was replaced by the Welsh Index of Multiple Deprivation 2005. This new index was produced in house in a partnership between the Welsh Assembly Government and the Local Government Data Unit ~Wales. The index was based on a similar methodology to the previous index and was more transparent and easier to understand (so that users could understand how the values were derived). It was based on a new geography called Super Output Areas (SOAs).

Super Output Areas

The Office for National Statistics developed a new geographic hierarchy called Super Output Areas (SOAs). They were designed to improve the reporting of small area statistics in England and Wales. Their first statistical application was for the Index of Deprivation for England in 2004, which led to them being widely used within local government. They have been increasingly used for data on the Neighbourhood Statistics (NeSS) website. It is anticipated that they will eventually become a standard for the production of National Statistics and will be used more generally.

Electoral divisions (previously known as wards) were the basic geographical units used for the Welsh Index of Multiple Deprivation 2000, but there were disadvantages with this approach. Electoral divisions vary greatly in size, from around 1,000 people to 20,000 (in Wales). This is not ideal for making comparisons throughout Wales, and it also means that data which can safely be released for larger electoral divisions may not be released for smaller ones due to disclosure rules (that is, the need to protect the confidentiality of individuals).

The boundaries of electoral divisions change. This creates problems when trying to compare data from different time periods. ONS decided to develop a range of areas that would be of consistent size and whose boundaries would not change. These would be built from groups of the Output Areas (OAs) used for the 2001 Census, and would be known as Super Output Areas (SOAs).

There are three layers of SOAs: Lower Layer, Middle Layer, and Upper Layer. This was because disclosure requirements mean that some sets of data could be released for much smaller areas than others. So to support a range of potential data requirements it was decided to create these three SOA layers. These are the constraints:

- A Lower Layer SOA must have a minimum population of 1,000. The mean size of all the Lower Layer SOAs must be close to 1,500. They are built from groups of Census OAs (usually between four and six).
- A Middle Layer SOA must have a minimum population of 5,000. The mean size of all the Middle Layer SOAs must be close to 7,200

- An Upper Layer SOA must have a minimum population of about 25,000: the formal definition has not been finalised.

There are 1,896 Lower Layer SOAs in Wales (34,378 in England and Wales). They were generated by a computer program which merged OAs taking into account population size, mutual proximity, and social homogeneity. The boundaries were released in February 2004.

The following table gives the number of Lower Layer SOAs in each local authority in Wales.

Number of LSOAs by local authority

	number of Lower Layer SOAs		number of Lower Layer SOAs
Isle of Anglesey	44	Neath Port Talbot	91
Gwynedd	75	Bridgend	85
Conwy	71	The Vale of Glamorgan	78
Denbighshire	58	Cardiff	203
Flintshire	92	Rhondda, Cynon, Taff	152
Wrexham	85	Merthyr Tydfil	36
Powys	80	Caerphilly	110
Ceredigion	47	Blaenau Gwent	47
Pembrokeshire	71	Torfaen	60
Carmarthenshire	112	Monmouthshire	58
Swansea	147	Newport	94

There are 413 Middle Layer SOAs Wales (7,193 in England and Wales). They were generated in two stages: a draft set was generated automatically (like the Lower Layer SOAs), then local authorities and other local agencies were invited to propose changes to the draft boundaries in order to establish areas that better met local needs. The consultation ran from March to May 2004 and led to publication of the final Middle Layer SOAs in August 2004.

The Upper Layer SOAs have not yet been formally defined.

Further information

The reports and data from WIMD 2008, together with guidance and Frequently Asked Question plus background to the project can be found on the WIMD 2008 web pages:

http://wales.gov.uk/topics/statistics/theme/wimd/?lang=en

2 Income Domain

Introduction

The purpose of this domain is to capture the extent of deprivation relating to income at a small area level across Wales. It focuses on the proportion of people living in households with income below a defined threshold or claiming benefits relating to low incomes.

The indicators used for the income domain were:

Adults & Children in Income Support households
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.

Income Support is intended to help people on low incomes who do not have to be available for employment. It can normally be claimed by people who are: aged 16 - 60; working less than 16 hours a week (and/or with a partner working less than 24 hours a week); not required to be available for full-time employment; and in receipt of insufficient income to meet prescribed needs. The main beneficiary groups are lone parents, the long- and short-term sick, people with disabilities and other special groups.

Adults & Children in Pension Credit households
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.

Pension credit is intended to help people on low incomes who do not have to be available for employment. It can normally be claimed by people who are: aged 60 or over; working less than 16 hours a week (and/or with a partner working less than 24 hours a week); not required to be available for full-time employment; and in receipt of insufficient income to meet prescribed needs. The main beneficiary groups are the long- and short-term sick and people with disabilities.

Adults & Children in Income-Based Job Seekers Allowance households
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.

This indicator captures all persons living in a household dependent on Income-based Job Seekers Allowance (JSA). JSA can be claimed by people aged 16 and over who are available for and actively seeking employment, including those in remunerative work for less than 16 hours a week on average, and by people on a government training scheme.

Adults & Children in Tax Credit (Child Tax Credit and Working Tax Credit) households below a low income threshold
Overall count by LSOA for August 2005 (2005-06 finalised awards) of in-work claimants of CTC and the out of work not claiming IS/JSA with equivalised income less than 60 per cent of the Wales median supplied by HMRC.

This indicator captures all persons living in a household dependent on the following Tax Credits:

1) Child Tax Credit – an income-based benefit for low-income families who have responsibility for a child(ren) under 16, or under 20 if in full-time non-advanced education or approved training.
2) Working Tax Credit – an income-based benefit for working adults available to households with adults in one of four categories:
   Work 16 hours or more a week and meet one of the following criteria:
   a) have responsibility for a child;
   b) have a disability that puts you at a disadvantage in getting a job;
   c) qualify for a 50-plus element;
   OR
   Work 30 hours or more a week and are 25 or over.

The methodology used for the selection of Tax Credit recipients in the income domain restricts inclusion to those below 60 per cent of the national median income, excluding housing benefit and before housing costs. Tax credits are included in WIMD 2008 as they provide an indication of families with low/moderate household incomes who are not on benefit.

National Asylum Support Service (NASS) supported asylum seekers in Wales in receipt of subsistence only and accommodation support
Count by LSOA of asylum seekers in receipt of subsistence only support and in dispersed accommodation in Wales as at the end of June 2007 supplied by the Home Office.

This indicator represents the presence of asylum seekers in Wales and contains people who are not eligible for the other benefits in the domain, yet are income deprived.

Calculating the Income Domain

Data for the benefits indicators above were provided by the Department for Work and Pensions (DWP) as an overall count by lower layer super output area (LSOA). By overall count this means that if an individual is claiming two or more of the benefits listed, they would only be counted once. A four quarter average was taken (using data for August 2006, November 2006, February 2007 and May 2007) to reduce seasonal variation.

Data on Tax Credits was provided by Her Majesty's Revenue and Customs (HMRC) as a count by LSOA of in-work claimants of Children's Tax Credit and the out of work not claiming IS/JSA, where equivalised income is less than 60 per cent of the national median income. Data is for 2005-2006.

Data on asylum seekers was taken for June 2007 and was obtained from the Home Office and the National Asylum Support Service (NASS) for all asylum seekers.

The denominator used for this domain was the 2006 SAPE population estimates (all ages), see Appendix B.

Summary of calculation methodology:

Use of shrinkage technique	Shrinkage was not applied
Combining indicators	Indicators were counts of unique individuals, or duplicates were removed, so that indicators could simply be summed and expressed as a percentage of the total population of the LSOA.
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

3 Employment Domain

Introduction

The purpose of this domain is to capture the extent of deprivation relating to employment at a small area level across Wales, and focuses on the proportion of working age people claiming out-of-work benefits.

The indicators used for the employment domain were:

Claimants of Unemployment-related benefits
Overall count of claimants by LSOA for the 12 moths from September 2006 to August 2007 obtained from the ONS (NOMIS).

This indicator captures all persons of working age (18-59 for women and 18-64 for men), who are claiming Job Seekers Allowance, available to those who are actively seeking work, but not in work.

Claimants of Incapacity Benefit/Severe Disablement Allowance
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.

Incapacity Benefit claimants are those of working age who are unable to work due to illness or disability and who meet criteria contribution conditions. All three rate categories of claimants are included (short-term lower, short-term higher and long term). Severe disablement Allowance is available to those people who are incapable of work and do not satisfy the contribution conditions of Incapacity Benefit. These two benefits were separately counted for WIMD 2005. However, due to the ever decreasing numbers of those receiving Severe Disablement Allowance (since 2001 there have been no new claimants), it was decided to combine the two groups.

Participants on New Deal for Young People and Intensive Activity Period (for New Deal 25+) not included in unemployment-related benefit counts
Overall count by LSOA from 2005 supplied by DWP.

This indicator captures young people (18-24) who have been claiming JSA for at least 6 months and persons 25+ who have been claiming JSA for at least 2 years and have therefore moved into the New Deal programme. This indicator was chosen as it captures people who are on a scheme which encourages them back to work, but who have not found employment yet and are not picked up on the JSA count.

Participants on New Deal for Lone Parents
Overall count by LSOA from 2005 supplied by DWP.

This is a new indicator, included to incorporate a large and statistically important group, not included in other indicators for this domain.

The only change made to this domain from WIMD 2005 was the inclusion of participants on New Deal for Lone Parents.

Calculating the Employment Domain

Data for three of the indicators above (Incapacity Benefit, Severe Disablement Allowance and New Deal participants) were provided by the Department for Work and Pensions (DWP) as an overall count of all four benefits at LSOA level. A four quarter average was taken (using data for August 2006, November 2006, February 2007 and May 2007) for the benefits data and the New Deal data used was from 2005. Counts of claimants of unemployment related benefits were obtained from the ONS (NOMIS) at LSOA level for the period September 2006 to August 2007.

The denominator used for this indicator was the SAPE 2006 LSOA population estimates for those of working age (18-59 for women and 18-64 for men).

Summary of calculation methodology:

Use of shrinkage technique	Shrinkage was not applied
Combining indicators	Indicators were counts of unique individuals, or duplicates were removed, so that indicators could simply be summed and expressed as a percentage of the working age population of the LSOA.
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

4 Health Domain

Introduction

The purpose of this domain is to capture the degree to which people are deprived of good health.

The indicators used for the health domain were:

Limiting long-term illness (LLTI)
Data supplied by the ONS for all people usually resident in the area at the time of the 2001 Census who had a limiting long-term illness. A limiting long-term illness covers any long-term illness, health problem or disability that limits daily activities or work.

The question in the Census on limiting long-term illness is well established and has been shown by numerous studies to be a valid measure of morbidity in the community.

Standardised all-cause death rate
Counts of all-cause mortality by LSOA with 5-year age and sex breakdown for the 10 year period 1997-2006 supplied by the ONS. Data was age-sex standardised using direct standardisation and a 6-year average (2001 to 2006) of ONS population estimates.

Poor health manifests itself both through a poorer quality of live but also in lower life expectancy which can be captured through age and sex standardised death rates.

Standardised cancer incidence rate
Count of all cases of cancer (all malignancies excluding non melanoma skin cancer) at LSOA level by sex, five year age band for the ten year period 1996-2005 supplied by Velindre NHS Trust.

A number of health conditions have strong links with deprivation (e.g. heart disease, cancer incidence etc.). Cancer incidence data are robust both in terms of the long standing central collection of the data and the numbers of cases involved even at a small area level.

Singleton low birth weights
Percentage of live singleton babies classed as having a low birth weight (defined as birth weight less than 2500g) using data for the 10 year period 1997 to 2006 supplied by the ONS.

Evidence suggests that low birth rate is linked to the mother's lifestyle and health. Low birth rate can also cause problems for a baby in later life increasing the risk of chronic diseases.

The new indicator, singleton low birth weights, was proposed for use in WIMD 2005, however at a late stage quality issues emerged with the allocation of data to small areas and with matching known totals. For WIMD 2008 the data was more robust and issues had been resolved, enabling it to be included in the health domain.

Calculating the Health Domain

Data for the first three indicators listed above were obtained for quinary age band and sex, up to age 84 and then a single age band of 85+ by sex. LLTI data was taken from the 2001 Census. Deaths data was provided by the ONS for the 10-year period 1997 to 2006. Cancer incidence data was provided by Velindre NHS Trust for the 10-year period 1996 to 2005. Singleton low birth weight data was provided by the ONS for the 10-year period 1997 to 2006.

Shrinkage was applied to all the indicators for WIMD 2005, however shrinkage was not applied to any of the indicators for WIMD 2008. Further analysis showed that removing shrinkage had little effect on the results and where there was an effect the unshrunk data provided a more sensible outcome.

The data were age-sex standardised using direct standardisation. Standardisation is needed to remove difference between areas that are purely down to different demographic profiles, e.g. a greater proportion of older people. Direct standardisation applies the overall Wales population distribution to the actual number of cases (whether it is LLTI, death or cancer incidence) in each area. In other words, the rate (e.g. death rate) for a particular age and sex group is applied to the Welsh population figure in that group. The sum of all these standardised age-sex group figures is divided by the overall Wales population to calculate the standardised rate.

The populations used were 2001 Census populations for the LLTI indicator and the death rate, cancer incidence and low birth weight indicators used a 6-year average (2001 to 2006) using SAPE population estimates.

Summary of calculation methodology:

Use of shrinkage technique	For WIMD 2005, shrinkage was applied to all indicators. However, shrinkage was not applied for any of the indicators for WIMD 2008.
Combining indicators	Indicators were combined using factor analysis, see Appendix C. The resultant weights for each indicator was as follows: 0.22 Limiting long-term illness 0.48 All cause death rate 0.19 Cancer incidence 0.10 Low Birth Weight
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

5 Education Domain

Introduction

The purpose of this domain is to capture the extent of deprivation relating to education, training and skills at a small-area level across Wales. It is designed to reflect the 'stock' and 'flow' or educational disadvantage within an area, by capturing low attainment among children and young people and the lack of qualifications and skills in adults.

The indicators used for the education domain were:

Key Stage 2, average point scores
Average point scores of pupils being taught in National Curriculum year group 6 in all schools (excluding Independent schools) in 2005, 2006 and 2007, measured using teacher assessments. Data is used from PLASC and postcodes matched to LSOAs using a look-up supplied by the Welsh Assembly Government Cartographics Unit.

3-years worth of data was used to reduce the impact of small numbers at the LSOA level. It is measured using teacher assessments, unlike the index in England which uses test scores. It should be noted that teacher assessments and tests are highly correlated. Pupils attaining level 'N' were excluded from the calculations because a grading of N denotes no information available so no accurate point score could be applied to such pupils. Independent schools were excluded for WIMD 2008, because it is not statutory for them to submit KS2 information to the Welsh Assembly Government and also as they do not complete PLASC there is no information available at postcode level for individual pupils. Data are collected annually by the Welsh Assembly Government.

Level	Points
Disapplied[1]	0
Absent[2]	0
Working towards level 1	3
1	9
2	15
3	21
4	27
5	33
6	39
7	45
8	51
Exceptional Performance	57

1. Children can be disapplied from following the National Curriculum, mainly as a result of them having a statement of special educational needs prepared by their LEA. Children can also be temporarily disapplied if they have recently arrived from a different educational system, had spells in hospital, been educated from school or been excluded. Children can be disapplied from any or all subjects.

2. Using Key Stage Teacher Assessments eliminates the issue of children who were absent from exams as every pupil receives and assessment score based on their overall year performance. Pupils can only be awarded an Absent mark if, in the opinion of the school, they have missed a sufficient proportion of the year that it would be inappropriate to award a level.

Key Stage 3, average point scores

Average point scores of pupils being taught in National Curriculum year group 9 in all schools (excluding Independent schools) in 2005, 2006 and 2007, measured using teacher assessments. Data is used from PLASC and postcodes matched to LSOAs using a look-up supplied by the Welsh Assembly Government Cartographics Unit.

3-years worth of data was used to reduce the impact of small numbers at the LSOA level. See the description of the KS2 indicator above for further details, including the points scores used. Data are collected annually by the Welsh Assembly Government.

Key Stage 4, average point scores

Average point scores of pupils aged 15 as at 31 August 2005 and 2006 in all schools. All approved GCSE and GNVQ results are included for 2005 and 2006. Data is used from PLASC and postcodes matched to LSOAs using a look-up supplied by the Welsh Assembly Government Cartographics Unit.

Points were allocated on the following basis:

GCSE/GCSE short course

Grade	GCSE score	GCSE short course score
A*	8	4
A	7	3.5
B	6	3
C	5	2.5
D	4	2
E	3	1.5
F	2	1
G	1	0.5

GNVQ

Level	Grade	Full GNVQ points	Part 1 GNVQ points
Intermediate	Distinction	30	15
	Merit	24	12
	Pass	20	10
Foundation	Distinction	16	8
	Merit	12	6
	Pass	6	3

Data are collected annually by the Welsh Assembly Government.

Primary school absence rates
Average pupil proportions of total half-day sessions absence (including both authorised and unauthorised absence). Data on the number of school sessions missed due to authorised and unauthorised absence is collected from the start of the academic year up to the date of the late May bank holiday. The indicator used the total absences, for the 2005/06 and 2006/07 academic years, taken from PLASC.

This new indicator provides a balance with the Secondary school absence rates previously included.

Secondary school absence rates
Average pupil proportions of total half-day sessions absence (including both authorised and unauthorised absence). Data on the number of school sessions missed due to authorised and unauthorised absence is collected from the start of the academic year up to the date of the late May bank holiday. The indicator used the total absences, for the 2004/05, 2005/06 and 2006/07 academic years, taken from PLASC. Independent schools were not included due to unavailability of data for all schools.

Proportion of people not entering Higher Education aged 18-19
The proportion of people not entering HE was then calculated using data on young participation rates for the 1998 to 2005 cohorts for Welsh LSOAs supplied by the Higher Education Funding Council for England (HEFCE).

This is a new indicator for WIMD 2008, formerly part of a combined people aged 16-18 (inclusive) in further or higher education in WIMD 2005.

Number of adults aged 25 – 59/64 with no qualifications
This indicator captures all adults aged 25 to retirement age (59 for women and 64 for men) who reported at the time of the 2001 Census as holding no qualifications or qualifications which do not reach the standard to be categorised as Level 1 qualifications.

These persons were chosen because persons with no qualifications are less likely to have functional literacy and numeracy skills, more likely to be economically inactive and on average earn less.

The WIMD 2005 indicator on people aged 16-18 (inclusive) in further or higher education was split into two separate indicators for WIMD 2008. They were: Proportion of people aged 16-17 not in Further Education and Proportion of people not entering Higher Education aged 18-19. Non-continuation following compulsory school age may be a reflection of a lack of opportunities. Also it will impact on attainment in higher level qualifications and there is a link between qualification levels in adult life, economic activity and salary. There is also a likely link between entry into HE and examination results in schools.

Data was collected and LSOA rates were calculated for the Further Education indicator. When this and the other indicators were analysed using Factor Analysis it was found that the Further Education indicator had a very small

and negative weight, suggesting that it was not a valid indicator for this domain. Further analysis showed that the amount of variance in the data explained by the indicators hardly changed when the Further Education indicator was taken out of the calculations. Because of this, the Further Education indicator was not included in the domain.

Calculating the Education Domain

Key stage 2 and Key stage 3 average point scores
Point scores for English, Welsh, Mathematics and Science were obtained by from the National Data Collection KS2 and KS3 databases. Because not all children are assessed in Welsh as a first language at key stages 2 and 3, the highest score in English and Welsh was taken along with the score in Mathematics and Science to provide comparability across Wales. Pupil-level data were then aggregated to postcode to give a count of pupils and a total point score. Postcodes were then matched to a LSOA using the apportionment method described in Appendix A.

When the data was allocated to LSOAs 10,712 pupils across the 3 Key Stage exams (KS2, KS3 and KS4) could not be matched to an LSOA. To reduce the number of unmatched data records a separate matching exercise was undertaken to identify non-matched postcodes that may have been inputted incorrectly. This was completed by looking for similar postcodes (i.e. those with one number or letter either side on a computer keypad) in the school catchment area. 5,117 pupils out of 105,385 Welsh domiciled pupils undertaking KS2 exams could not be matched to a LSOA. For KS3 exams, 3,229 pupils out of 112,238 Welsh-domiciled pupils undertaking KS3 exams could not be matched to a LSOA.

For each LSOA an average point score was calculated for all pupils in that LSOA and as such no separate population denominator was needed.

Key stage 4 average point scores
Pupil numbers and total point score by postcode were obtained from the school performance database and the public level annual school census (PLASC). In line with usual practice in Wales, all exams taken by pupils aged 15 were included in the point score.

Postcodes were then matched to a LSOA using the apportionment method described in Appendix A. After the data was allocated to LSOAs, in order to reduce the number of unmatched data records a separate matching exercise was undertaken to identify non-matched postcodes that may have been inputted incorrectly. This was completed by looking for similar postcodes (i.e. those with one number or letter either side on a computer keypad) in the school catchment area. 2,366 pupils out of 73,742 Welsh domiciled pupils undertaking KS4 exams could not be matched to a LSOA.

For each LSOA an average point score was calculated for all pupils in that LSOA and as such no separate population denominator was needed.

Primary school absence rates
School-level primary school absence data are collected annually for the total number of half-day sessions missed and the total number of half-day sessions within the

academic year. As no lower level data were available, LSOA figures had to be constructed from this school level data. However, pupils living in a given LSOA could come from different schools. To account for this the following approach was adopted:

- total primary school pupils of compulsory school age by school and postcode were obtained from PLASC;
- all pupils from a particular school were given the school average number of half-day sessions missed and total half-day sessions (i.e. the average being the total number for the school divided by the total number of pupils at the school);
- all postcodes were then allocated to a LSOA using the apportionment method described in Appendix A;
- the number of half-day sessions missed and the total number of half-day sessions were then aggregated for all pupils in a LSOA (regardless of school);
- the absence rate for the LSOA was calculated as the total of half-day sessions missed as a proportion of total half-day sessions.

Secondary school absence rates

School-level secondary school absence data are collected annually for the total number of half-day sessions missed and the total number of half-day sessions within the academic year. As no lower level data were available, LSOA figures had to be constructed from this school level data. However, pupils living in a given LSOA could come from different schools. To account for this the following approach was adopted:

- total secondary school pupils of compulsory school age by school and postcode were obtained from PLASC;
- all pupils from a particular school were given the school average number of half-day sessions missed and total half-day sessions (i.e. the average being the total number for the school divided by the total number of pupils at the school);
- all postcodes were then allocated to a LSOA using the apportionment method described in Appendix A;
- the number of half-day sessions missed and the total number of half-day sessions were then aggregated for all pupils in a LSOA (regardless of school);
- the absence rate for the LSOA was calculated as the total of half-day sessions missed as a proportion of total half-day sessions.

Proportion of people not entering Higher Education aged 18-19

Data on young participation rates for the 1998 to 2005 cohorts for Welsh LSOAs were supplied by the Higher Education Funding Council for England (HEFCE). The proportion of people not entering HE was then calculated for each LSOA as 1 less the participation rate.

Proportion of Adults with No or Low Qualifications

LSOA-level data from the 2001 Census was obtained from the Office for National Statistics for females aged 25-59 and males aged 25-64 with no qualifications. This also incorporates qualifications that do not meet the criteria to be classified as Level 1 by the Department for Education and Skills National Qualification Framework.

The lower age bound of 25 was used to reduce the effect on areas which are highly populated with college students (who are well qualified), who stay only on a temporary basis, hiding the true level of qualifications permanent residents hold within an area. The denominator used for this indicator was the 2001 Census count of all females aged 25-59 and all males aged 25-64 within the LSOA.

Summary of calculation methodology:

Use of shrinkage technique	For WIMD 2005, shrinkage was applied to all indicators except secondary school absences. However, shrinkage was not applied for any of the indicators for WIMD 2008.
Combining indicators	Indicators were combined using factor analysis, see Appendix C. The resultant weights for each indicator was as follows: 0.08 Key Stage 2 average point scores 0.25 Key Stage 3 average point scores 0.16 Key Stage 4 average point scores 0.05 Primary school absence rates 0.04 Secondary school absence rat 0.24 Proportion of people not entering Higher Education aged 18-19 0.18 Proportion of adults with no or low qualifications
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Issues highlighted in the calculation of the domain

Cross border student flows
Some Welsh-domiciled students attend schools in England and are therefore not included in data obtained from Welsh schools. To overcome this, data were requested from DfES for every Welsh domiciled student attending an English school on the same scoring system as used for Welsh schools. Similarly, English-domiciled pupils attending Welsh schools were removed from the data (i.e. those pupils with postcodes in England).

Small numbers
For the WIMD 2008, 3-years worth of data was used for the Key Stage 2 and 3 indicators and 2-years worth of data was used for the Key Stage 4 indicator. Using a number of years data for each Key Stage helped overcome the problem of small

numbers, rather than allowing the shrinkage technique to adjust values based on a few pupils toward the local authority average as was done in WIMD 2005.

Shrinkage
For WIMD 2005 shrinkage was used. Subsequently in a review of the Scottish Index of Multiple Deprivation, it was recommended that shrinkage should no longer be used. Tests on Welsh data have shown that not using shrinkage has no significant impact on the groupings of the LSOAs e.g. the vast majority top 10% LSOAs stay within the top 10% within the domain.

Quality Assurance
As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

6 Housing Domain

Introduction

The purpose of this domain is to capture deprivation though a lack of adequate housing.

The indicators used for the housing domain were:

Lack of central heating
The number of people living in households without central heating as a proportion of all people in households, derived from the 2001 Census.

This is a proxy measure for quality of housing and is used in preference over more direct measures as it is available at an LSOA level.

Overcrowding (excluding all student households)
All people in households, excluding people living in all student households, living in households with an occupancy rating of -1 or less as a proportion of all people in households excluding people living in all student households, derived from the 2001 Census.

This is a measure of personal circumstance and is a reflection of the inability to access suitable accommodation, either through low income or the availability of housing. The measure used is the 2001 Census "occupancy rating", where households are overcrowded on a value of -1 or less. This is defined as having one or more rooms too few for the size of household.

The WIMD 2008 housing domain is the same as used for WIMD 2005. Work to develop availability of small area data is ongoing. Potential indicators for use in future updates of WIMD include lack of demand, physical condition and social housing.

Calculating the Housing Domain

Both indicators were derived from the 2001 Census and rates were derived using Census denominators. In both cases, data for households (i.e. excluding communal establishments) was used and in addition all student households were excluded for the overcrowding indicator.

Summary of calculation methodology:

Use of shrinkage technique	Shrinkage was not applied
Combining indicators	Factor analysis could not be applied with only two indicators and so an equal weighting of the normally transformed ranks of the indicators was used (i.e. 50/50).
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

7 Physical Environment Domain

Introduction

The purpose of this domain is to measure environmental factors that may impact on quality of life in an area. Negative impacts on quality of life are judged to be a form of deprivation. Environmental deprivation is generally not correlated with social or economic deprivation in Wales. (ref: Walker et al 2003). This domain does not capture aspects of deprivation such as health inequalities (this is an objective of the Health domain). Data covering Wales is not available on actual impacts on quality of life; so factors were chosen that were judged to indicate an increased potential for reduced quality of life and had data available.

The indicators used for the physical environment domain were:

Air Quality
Air pollution concentrations are low-resolution data estimated using models. They are estimated averages over an area. Concentration maps are validated against available monitoring data. The latest air quality maps were for 2006 and data was supplied by the Environment Agency.

Air quality is believed to be a good proxy measure of the quality of the surrounding environment. Poor air quality suggests proximity to certain activities such as traffic, domestic combustion and industrial sites – activities that could have a negative impact on quality of life, the local environment and health.

The mapped data for each pollutant, together with Air Quality Management Area data, was normalised using air quality standards and exponentially transformed to create the indicator in one combined map. This helped overcome some of the issues encountered in WIMD 2005, such as when using an average measure of all pollutants, information on single pollutants at high concentrations could be lost in the deprivation score.

Air Emissions
Air pollution emissions are low-resolution data estimated using models and emissions inventories. They are estimated averages over an area. Concentration maps are validated against available monitoring data. The latest emission maps were for 2005 and data was supplied by the Environment Agency.

While it is accepted that air quality is the preferred measure of risks from air pollution, air emissions data provides a good set of complimentary data covering pollutants not included in the Air Quality indicator. Emissions data are good indicators of proximity to polluting activities. Mapping of air emissions data was carried out in the same way as for air quality data.

Flood risk
Flood risk calculations and maps were based on the 2007 NAFRA database which includes flood risk, taking into account flood defences where these are known. Data was supplied by the Environment Agency.

This indicator was used in the WIMD 2005, under the name "proportion of people living in an area with a significant risk of flooding". Those who suffer flooding have a significant lowering of quality of life that can last for a number of years. Homes in areas that suffer increased flood risk will often have significantly higher insurance premiums, potentially leading to higher financial hardship in these areas. It is likely that economically and socially deprived areas will take longer to recover from flood events.

For WIMD 2008, different levels of risk were taken into account, as is done with insurance companies with 3 levels of risk; significant, moderate and low risk. The previous problem that existed was the lack of flood defence information in the data and that WIMD would need to be sensitive to these interventions. In consequences, The Environment Agency now has risk maps including most flood defences. The risk is based on frequency rather than level of damage caused by any flooding.

Proximity to waste disposal and industrial sites
Each site was assigned a proximity score which relates to the potential and actual deprivation the site could cause to the environment and the people living in its vicinity. The OPRA score was used to assign the score for those sites where it is currently recorded. Where no OPRA score was available a proximity score was created based on best available information. Where no boundary data was available, the boundary of a site was assumed to be circular with a radius of 0.5km. The data is from 2007 and was supplied by the Environment Agency.

The proximity to waste disposal and industrial sites indicator was used in WIMD 2005, under the two separate indicators proportion of residential population living within 1km from current and recent waste disposal sites (landfills and incinerators), and proportion of residential population living within 1km from a significant industrial source (those identified in Part A(1) of The Pollution Prevention and Control Regulations 2000). The indicator was chosen because some communities and individuals view close proximity of waste disposal and industrial sites as undesirable.

Some sites have a greater potential to impact on quality of life than others – a direct quantification of this is not available. In a change from WIMD 2005, WIMD 2008 classified sites in order to define different buffer zones according to hazard assessments. This was done using a more scientific and robust process, which took into account process type and size and operator performance.

It was hoped to include an indicator on Ground Movement Risk due to it being a major risk to building foundations. However, the indicator was not included in the domain as it was not possible to define the indicator in such a way that it was compatible with the Flood Risk indicator already in the domain.

Calculating the Physical Environment Domain

In all indicators a measure of actual population location was required and postal address points were used. To reflect the fact that individual householders are the focus of WIMD, the following adjustments were made to the Royal Mail 2005 address point file:

Exclude:
- points outside of Wales
- PO Boxes
- addresses containing
  Organisation Name (ON)
  Department Name (DP)

but include if organisation name:
- ends in FARM
- indicates HOSPITAL
- indicates HOME (Residential, Care, Nursing Care, Rest)
- indicates MOD Barracks

Air Quality
A 1km x 1km vector (polygon) grid was generated to cover Wales, corresponding to concentration GRIDs supplied by Netcen. Concentration values were then extracted for each cell for the following pollutants and statistics:

Benzene annual mean
Carbon Monoxide annual mean
Ozone maximum daily 8 hour mean
Nitrogen Dioxide annual mean
Particulates annual mean
Sulphur Dioxide annual mean
99.9th percentile of Sulphur Dioxide 15-minute means

For land-based cells without data, concentration values were inferred using a simple average of surrounding cells.

Figures for LSOAs were calculated by examining the overlap of the 1 km grid data with each LSOA and averaging the results of each grid intersecting the LSOA, weighted by the number of address points in the intersection. Each pollutant value was then adjusted to an equivalent scale using a factor based on the objective, standard or risk factor for that pollutant and statistic.

Air Emissions
A 0.5km x 0.5km vector (polygon) grid was generated to cover Wales, corresponding to concentration GRIDs supplied by Netcen. Annual emission values were then extracted for each grid cell for the following pollutants:

Arsenic
Benzo[a]pyrene
Butadiene
Cadmium

Chromium
Dioxins
Mercury
Ammonia
Nickel
Nitrogen Oxides
Lead
Vanadium
Volatile Organic Chlorides

Figures for LSOAs were calculated by examining the overlap of the 0.5 km grid data with each LSOA and averaging the results of each grid intersecting the LSOA, weighted by the number of address points in the intersection. Each pollutant value was then adjusted to an equivalent scale using a factor based on objective, standard or risk factor for that pollutant.

Flood risk
Using a Geographical Information System (GIS), the overlap of flood risk areas (Fluvial, Fluvial/Tidal, Tidal) and LSOAs was calculated. The Environment Agency now has risk maps including most flood defences and has developed a way of mapping flood risk based on an analysis of the proportion of households in each flood risk zone. Mapping is updated on a 3 monthly basis, covering main rivers and coasts. The risk is based on frequency rather than level of damage caused by flooding.

Proximity to waste disposal and industrial sites
In a change from WIMD 2005, WIMD 2008 classified sites in order to define different buffer zones according to hazard assessments. This was done using a more scientific and robust process, which took into account process type and size and operator performance.

Where available the Environment Agency 'Environment Protection Operator and Pollution Risk Appraisal' (EPOPRA*) scores were used as a basis for categorising waste disposal and industrial sites. Integrated Pollution Prevention and Control Directive (IPPC) sites, A1and A2 processes sites were included, along with Control of Major Accident Hazards (COMAH) sites, nuclear sites, waste management sites and Sewage treatment works. Many of these types of sites do not have an EPOPRA score. Where there is not an EPOPRA score, a proxy score was allocated by the Environment Agency relating to size and processes and buffer zones calculated accordingly. The Proximity to Waste Disposal and Industrial sites indicator was agreed as an indicator, which should relate sites with an EPOPRA or proxy score to where people live. If more people live in close proximity to an industrial or waste site (which would be defined by a buffer) then the greater the level of deprivation. The banding used detailed below allocates a greater weighting and greater number of buffers (greater distance effect) to sites in a higher band. Those in band '0' are judged to have no significant effect (weighting = 0); those in band '5' are judged to have the greatest effect (weighting varies from 5 to 0 with distance); If a household is within two buffers (overlapping) then the proximity score will be counted twice accordingly.

1. Assign each site into one of 6 bands, depending on the EPOPRA score where available.

2. Use the bands to identify the size and weighting of the buffers to apply around each site using the weighting matrix:

Distance from site (km)	Band created from EPOPRA score (Industrial and waste management sites)					
	5	**4**	**3**	**2**	**1**	**0**
<=0.5	5	4	3	2	1	0
>0.5 to 1	4	3	2	1	0	0
>1 to 1.5	3	2	1	0	0	0
>1.5 to 2	2	1	0	0	0	0
>2 to 3	1	0	0	0	0	0
>3	0	0	0	0	0	0

Sites without EPOPRA scores: Nuclear sites - Band 4, A2 sites - Band 2 (affect up to 1 km), STWs - Base on population. Largest population in Band 2

STW Population Equivalent	Band for profile
>100,000	2
>10,000 - 100,000	1
<= 10,000	0

* EPOPRA is a risk-screening tool that the Environment Agency use to regulate operators under the Integrated Pollution Prevention and Control (IPPC) regulations and the Waste Management Licensing (WML) regulations. It has been used for IPPC since 2003 and for WML since April 2005. IPC sites also have EPOPRA scores, which are derived in a simpler way. EPOPRA is used to help target regulatory effort at those activities that present the greatest risk to the environment.

Summary of calculation methodology:

Use of shrinkage technique	Shrinkage was not applied
Combining indicators	The domain was broken down into 3 sub-domain of like measures: • air pollution (air quality/emissions) • flood risk • proximity to regulated sites the first sub-domain having two indicators and the later two only a single indicator each. With less than three indicators in any one sub-domain, factor analysis could not be applied and so an equal weighting of the normally transformed ranks of the indicators within sub-domains was used. Sub-domains were then weighted together

	equally (one third each).
Exponential transformation	Before sub-domains were combined they were exponentially transformed. As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

As part of developing the domain originally for 2005, a comparative analysis using 2001 Census data at OA level and 2001 adjusted address points was undertaken to show that the use of adjusted address points alone reflected results taking into account population distributions. This was not done for 2008 since a more sophisticated method for combining the data later on in the process was used. Many quality assurance procedures were carried out on the GIS database before it was used in calculations.

Data Issues

- Shortage of appropriate data with a full coverage of Wales means that the domain can still be considered to be incomplete.
- Different sizes and types of regulated process will be associated with different levels and types of effect. In 2008 it was attempted to take this into account by assigning a proximity score which related to the potential and actual deprivation the site could cause to the environment and the people living in its vicinity. The OPRA score was used to assign the score for those sites where it is currently recorded. Where no OPRA score was available a proximity score was created based on best available information.
- Some non-regulated and illegal activities can have a significant local effect that may be greater than that of sites included in the proximity indicator.
- Where no boundary data was available, the boundary of a site was assumed to be circular with a radius of 0.5km.
- Air pollution concentrations and emissions are low-resolution data estimated using models and emissions inventories. They are estimated averages over an area. Concentration maps are validated against available monitoring data. The latest air quality maps were for 2006 and latest emission maps were for 2006.
- Flood risk calculations and maps were based on the 2007 NAFRA database which includes flood risk, taking into account flood defences where these are known. Flood defences reduce but do not eliminate flood risk. Local Authorities should seek further information from the Environment Agency flood risk maps before acting.

Interpretation Issues

Quantification and understanding of Environmental Deprivation and Environmental Injustice and how they link to quality of life are currently not well-developed subject areas. This together with data limitations will inevitably lead to problems in and debate on the interpretation of the domain.
Care should be taken in interpretation at smaller scales. Extrapolation below LSOA scale would not be possible. The domain is about areas in which people live and overall potential for deprivation for those communities.
Mapping the Environment domain indicators will not give maps of environmental quality or flood risk – the maps will identify LSOAs and the ranked proportions of their populations judged to be at risk of a reduced quality of life due to specific environmental factors.

8 Geographical Access to Services Domain

Introduction

The purpose of this domain is to illustrate the deprivation as a result of a household's inability to access a range of services, considered necessary for day-to-day living. The access is by walking or using public bus services or both.

The indicators used for the geographical access to services domain were:

Food shop
The average time taken for every household in a LSOA to travel to a food shop by walking or using public bus services or both.

This indicator is intended to cover the purchase of basic provisions (e.g. bread and milk). Service points include premises from the local corner shop up to large supermarkets.

GP surgery
The average time taken for every household in a LSOA to travel to a GP surgery by walking or using public bus services or both.

This indicator is intended to cover the day-to-day need for primary health care. This indicator includes all GP surgeries, although the services available across GP surgeries vary, the basic services are offered by all surgeries.

Primary school
The average time taken for every household in a LSOA to travel to a primary school by walking or using public bus services or both.

This indicator is designed to reflect the access to a primary school of children aged 4 to 11 to a primary school. Access is defined purely on the child's ability to access any primary school and takes no account of the school actually attended.

Post office
The average time taken for every household in a LSOA to travel to a post office by walking or using public bus services or both.

This indicator encapsulates accessibility to a post office. Post offices are used weekly by most of the community and are a vital communications source. This may include obtaining advice and assistance, sending and collection of mail/parcels, collection of benefits, payment of bills, withdrawal of money and many other services.

Public library
The average time taken for every household in a LSOA to travel to a public library by walking or using public bus services or both.

This indicator covers all static libraries and reflects a library's role as a vital modern communications service, with online access to further advice and

information, rather than just access to the more traditional services (e.g. book lending). For WIMD 2008 mobile libraries were also included.

Leisure centre
The average time taken for every household in a LSOA to travel to a leisure centre by walking or using public bus services or both.

This indicator is intended to cover an individual's ability to access facilities important for health and well being.

NHS dentist
The average time taken for every household in a LSOA to travel to a NHS dentist by walking or using public bus services or both.

This indicator is based on all dentists offering NHS treatment. It looks solely at an individual's ability to access a surgery and does not take into account whether spaces are available, i.e. the indicator is measuring purely geographical access.

Secondary school
The average time taken for every household in a LSOA to travel to a secondary school by walking or using public bus services or both.

This indicator is designed to reflect the access needs of children aged 11 to 16 to a secondary school. Access is defined purely on the child's ability to access any secondary school and takes no account of the school actually attended.

Transport Nodes
The average time taken for every household in a LSOA to travel to a transport node by walking or using public bus services or both.

This is a new indicator for WIMD 2008 to show access to long distance transport services. The proximity of transport nodes to each household was computed for coaches and rail routes.

For WIMD 2005 not only was the road network used for travel distances but public transport or walking or both was used to calculate the travel times. The reason for this is that access to cars is not just dependent on car ownership but the availability for a particular individual at a particular time of day and this cannot be modelled accurately. For WIMD 2008, average times were used to calculate the indicator values, as opposed to threshold times. While threshold times produce a sensible distribution of times by address point, when aggregated to a LSOA level the indicators do not then perform in the same way and do not provide a suitable distribution for the Index. Taking the average times for each LSOA to each service gives a much better LSOA distribution for the purpose of the Index.

Calculating the Access to Services Domain

The grid reference of service point locations were provided by the Cartographics Unit of the Welsh Assembly Government Statistics Directorate. Royal Mail address

points were used as the origin points for journeys individuals would need to make to these service points. Commercial address points were excluded in urban areas but included in rural areas to ensure farms were not excluded (not all of which have farm in the description in the address point file).

Calculations were undertaken in a package developed for local transport planning. The road network in Wales as well as bus routes and timetables were input to the package. The time needed to get from each origin to each service point was then calculated based on the distance and availability of buses. A maximum walk of 800 metres was set for the start and end parts of the journey, i.e. from home to the bus stop and from the bus stop to the service, or direct to the service point if that should apply. Up to 10 of the shortest trips were recorded from each household to each service type – in excess of 10 million trips were recorded.

The indicators were then calculated as the time taken for every household in a LSOA to travel to each of the services identified in the domain using the average of the 10 shortest trips.

The Public Transport and Road Network data is from 2007. The Postal Address file contains 2007 Welsh Assembly Government data and 2008 Ordnance Survey data.

Summary of calculation methodology:

Use of shrinkage technique	Shrinkage was not applied.
Combining indicators	Indicators were combined using factor analysis, see Appendix C. The resultant weights for each indicator was as follows: 0.14 access to a food shop 0.18 access to GP 0.16 access to primary school 0.10 access to post office 0.09 access to library 0.06 access to leisure centre 0.12 access to NHS dentist 0.07 access to secondary school 0.07 access to transport nodes
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

9 Community Safety Domain

Introduction

The purpose of this domain is to reflect where people are deprived of a safe community in which to live. Safety includes levels of household and personal crime and quality of experience in public places compatible with access to ordinary work, leisure and social relationships.

The indicators used for the Community Safety domain were:

Police Force Recorded Crime
Four crime types which affect individuals or businesses were selected; violence, burglary, criminal damage and theft. The incidents (crimes) were recorded by the police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent) and collected as counts by LSOA. The incidents were located to the point at which they occurred and allocated to the appropriate LSOA. Data are from 2005-2006 and 2006-2007.

The incidents (crimes) were recorded by the police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent). They were of selected types which affect individuals or businesses; violence, burglary, criminal damage and theft. The criteria for selecting incidents by crime code is similar to that used in England, but a small number of additional codes were included. The additional codes produced only a small number of additional incidents. The incidents were located to the point at which they occurred and allocated to the appropriate lower super output area (LSOA).

Youth Offenders
The number of offenders (aged 10 to 17 years of age) dealt with by the 18 Youth Offending Teams within Wales, using the person's address to allocate to a LSOA. Care homes and institutions were excluded from the counts. Data are from 2005-2006 and 2006-2007.

Offenders often come from the most deprived section of society. Offenders at liberty is a measure to indicate issues relating to young people that indicate their need and the social problems they cause. This indicator complements the recorded crime data by including the offenders at liberty in an area.

Within Wales there are 18 Youth Offending Teams. The offenders dealt with by those teams are aged from 10 to 17 years of age. Where a person had more than one address during the period each addresses in was counted provided it was in a different LSOA. Where there were multiple offenders at a single address a weight of 2 was applied. Care homes and institutions were excluded from the counts.

Adult Offenders
The number of offenders (aged 18 and over) dealt with by the 4 Probation Service Areas (North Wales, Dyfed Powys, South Wales and Gwent) within Wales, using the person's address to allocate to a LSOA. Offenders residing in institutions were included in the count if a residential address was provided. Data are from 2005-2006 and 2006-2007.

As with youth offenders, adult offenders are often drawn from the most deprived section of society. Offender information complements crime records and was collected by the probation service. Within Wales there are four probation service areas (North Wales, Dyfed Powys, South Wales and Gwent). The offenders dealt with by those teams are aged 18 and over. Where a person had more than one address during period each addresses in was counted provided it was in a different LSOA. Where there were multiple offenders at a single address a weight of 2 was applied. Offenders residing in institutions were included in the count if a residential address was provided.

Fire Incidence
Incidents of primary fires with the addition of "derelict vehicle" fires were collected as counts by LSOA. Secondary "derelict vehicle" fires were only included if available at the LSOA level. Data for the Fire Incidence indicator is from 2005 and 2006.

Incidents requiring call out of fire and rescue services are related to deprivation[1] and more likely within disadvantaged groups. Primary fires[2] include "all fires in buildings, vehicles and outdoor structures or any fire involving casualties, rescues, or fires attended by five or more appliances". Secondary fires are "the majority of outdoor fires including grassland and refuse fires unless they involve casualties or rescues, property loss or five or more appliances attend". Primary fires, which relate better to property and people, with the addition of "derelict vehicle" fires that tend to occur in deprived areas, was used. The secondary "derelict vehicle" fires were only included if available at the LSOA level.

The community safety domain was introduced for the WIMD 2008. It was considered for WIMD 2005 but was not included due to data quality issues.

Calculating the Community Safety Domain

Data for the Indicators above were collected as stated above as counts by lower layer super output area (LSOA). In some cases it was not possible to assign the occurrence (crime or offender) to a LSOA, these occurrences were apportioned across the relevant larger geographical area. Data for the Police Force Recorded Crime, Youth Offenders and Adult Offenders indicator are from 2005-2006 and 2006-2007. Data for the Fire Incidence indicator is from 2005 and 2006.

The denominator used for the burglary category for the crime indicator was the total number of dwellings (from the 2001 Census) plus the total number of business addresses (using the 2006 count). The denominator used for the theft, violence and criminal damage categories for the crime indicator was the total resident population

[1] Reference in the introduction, page 5 at:
http://www.scotland.gov.uk/Resource/Doc/36496/0024964.pdf
[2] Definitions are found in the FDR1 (Fire Damage Report) Guidance document introduction section or page 115 of Fire Statistics UK 2005 at:
http://www.communities.gov.uk/pub/25/FireStatisticsUnitedKingdom2005_id1509025.pdf

(from SAPE 2005) plus the total non-resident workplace population (aged 16-74) less the total prison population (mid 2005).

The denominator used for the youth offenders indicator was total resident population aged 10-17 (from SAPE 2005) less the total prison population aged 10-17 (mid 2005). The denominator used for the adults offenders indicator was total resident population aged 18 and over (SAPE 2005) less the total prison population aged 18 and over (mid 2005). The denominator used for the fire incidence indicator was the total resident population (SAPE 2005).

Summary of calculation methodology:

Use of shrinkage technique	Shrinkage was not applied.
Combining indicators	Indicators were combined using factor analysis, see Appendix C. The resultant weights for each indicator was as follows: 0.21 Violence 0.06 Burglary 0.07 Theft 0.39 Criminal Damage 0.06 Youth Offenders 0.16 Adult Offenders 0.05 Fire Incidence
Exponential transformation	As with all domains, the final domain ranks were exponentially transformed, see Appendix D.

Quality Assurance

For police incident reports, a number of years' data was collected and the distribution of crime codes recorded was checked between years' data sets, to investigate whether there were any systematic changes in the way the codes were applied. In addition checks were made for operational changes leading to more incidents of any particular type being recorded. The distribution of police recorded incidents was mapped using GIS and the distribution viewed against population and land use. There was a check to ensure there was no clustering around police stations due to default codes being applied.

When looking at the youth offending and probation records, the locations at which those under supervision were reported were checked against residential locations. This ensured the locations given were not clustered outside residential areas.

As a double check on calculations, while operations were set up as SAS programs to allow repeated use, the results of each stage were checked independently in Excel.

Factor Analysis

Factor analysis for the domain produced 2 factors, which was not ideal for calculating the domain and could indicate the presence of sub-domains or some issues with the underlying data. Further correlation analysis indicated that there were no sub-domains. There was some positive correlation between all the indicators, but with original parameters this was not strong enough to ensure only 1 factor. However 1 factor appeared to explain the majority of the variance, indicating that it was the dominant factor.

Discussions with Oxford University showed that the correlations for the English IMD crime domain were very similar to those calculated for WIMD, indicating that the underlying data showed similar patterns. Following further investigations and discussions, a single factor was produced by altering (increasing) one of the parameters used in the factor analysis calculations (minimum eigenvalue). From this the weights were easily calculated and applied to the indicators for calculating the final domain.

Appendix A Allocation of Data to Lower Layer Super Output Areas

Background

LSOAs are still a relatively new geography and as such data are generally not allocated to an LSOA as part of the collection process. Therefore for WIMD 2008, a method of allocating other geographical level data to LSOAs had to be devised.

Different data sets vary in their level of geographic coding and so different approaches are required. Data sets fall into one of the following categories, shown in order of preference in terms of data quality:

1. data are geocoded and can be allocated to LSOAs exactly using a Graphic Information System (GIS);
2. data contain the full postal address and can be allocated exactly to LSOAs using ONS lookup tables;
3. data are coded with some other small area geography and these can be allocated to LSOAs in some way, although the matching will not be exact.

The following sections explain the approach taken for WIMD 2008 in each case.

1 Geocoded data

Geocoding is still relatively rare particularly for national-level data sets. In the case of WIMD 2008 the only geocoded information available was the service point location information (e.g. schools, post offices etc.) used in the Geographic Access to Services Domain.

2 Address Matching

This is the most accurate method for allocating data that has not been geocoded. The ONS have created a lookup to allocate each individual address to a LSOA which forms part of the toolkit for data providers of Neighbourhood Statistics (a package called MatchCode).

There are two main issues with this approach. The first is to do with address quality; addresses are often incomplete or partially incorrect (e.g. missing house number or incorrect postcode). There may then be a residual of records for which another approach is required (e.g. allocate on the basis of postcode). The second problem has to do with data confidentiality and legal restrictions which prevent the supply of full address information. The solution to the second problem is for the data provider to address match data and provide LSOA level counts and this was the case with benefits data that had already been coded for use in Neighbourhood Statistics.

3 Allocating small area data to LSOAs

The most common small area identifier is the postcode, and this was the building block used for many indicators in WIMD 2008.

Postcode level data

LSOA boundaries do not fit exactly with postcode boundaries. However, it was initially assumed that the fit would be good, that the occurrence of postcodes being split by LSOA boundaries would be minimal and that splits would generally result in the majority of a postcode clearly within a single LSOA. A detailed investigation found that there was a high proportion of split postcodes and that the occurrence of postcodes effectively split down the middle was not negligible particularly in urban areas where LSOA boundaries can go down the middle of streets.

The standard approach with postcode level data is to allocate data on a best fit basis using the Postcode Address File (PAF), where postcodes are allocated to the area in which most of the population lies. As noted above, the match of postcodes to LSOAs in Wales is not a close fit and this causes concern with something like WIMD as rates are calculated using denominators from a different source and areas are then ranked. An alternative method of apportioning postcode data to each of the LSOAs that lie within the area was developed.

Apportionment of split Postcodes

The only information generally available with post code level data is the count (e.g. the number of deaths) and the rate that is derived from this. The basic principle used was that the rate for the given indicator should apply equally across the whole postcode and that this should be preserved under any allocation methodology. This can be done by weighting the postcode rates according to the proportion of the postcode population sitting within the LSOAs, as shown below:

$$LSOA\,rate = \frac{\sum\left(PC_{in}\ popn \times PC\ rate\right)}{LSOA\ popn}$$

Where $PC_{in}popn$ = population of the postcode within the LSOA
PC rate = postcode rate
LSOA rate = LSOA rate
LSOA popn = LSOA population

When postcodes are not split this essentially gives the rate a weighting of 1 as the postcode population figures cancel:

$$LSOA\,rate = \frac{\sum\left(PC\,popn \times PC\,num / PC\,popn\right)}{LSOA\,popn}$$
$$= \frac{\sum PC\,num}{LSOA\,popn}$$
$$= \frac{LSOA\,num}{LSOA\,popn}$$

where

PCpopn = total population of the postcode

PCnum = number of individuals in the postcode

LSOAnum = the total number of individuals in the LSOA.

When postcodes are split, it has been assumed that

$$\frac{PC_{in}\,popn}{PC\,popn} = \frac{Address\,Points_{in}\,num}{Address\,Points_{tot}\,num} = PercentagePCinLSOA$$

where

AddressPoints$_{in}$*num* = number of postal address points in the postcode that are within the LSOA

AddressPoints$_{tot}$*num* = total number of postal address points in the postcode

PercentagePCinLSOA = estimated percentage of the postcode population within the LSOA

This says that the proportion of address points is equivalent to the proportion of the population, and hence assumes that population is distributed in the same way within each sub-area within the postcode. While it will not always be the case that the number of persons per household is the same in all parts of a postcode it will not generally be dramatically different. However, there is no realistic way to account for this (2011 Census data are too far out-of-date to be representative and are not available at a postcode level), and the WIMD Methodology Working Group agreed that the approach would provide more robust results than simple best fit.

Under this assumption, the LSOA rate is calculated as follows:

$$LSOA\,rate = \frac{\sum\left(PC_{in}\ popn \times PC\ rate\right)}{LSOA\ popn}$$

$$= \frac{\sum\left(\frac{PC_{in}\ popn \times PC\ num}{PC\ popn}\right)}{LSOA\ popn}$$

$$= \frac{\sum\left(PercentagePCinLSOA \times PC\ num\right)}{LSOA\ popn}$$

Which is equivalent to

$$LSOAnum = \sum\left(PercentagePCinLSOA \times PCnum\right)$$

Hence, approximate counts for indicators can be constructed at LSOA level by taking the relevant percentage of the counts for each postcode that falls within it. While this is the most straightforward way to estimate an LSOA figure, the above analysis shows that it is actually equivalent to giving every individual in split postcodes the rate for that postcode.

To facilitate this approach the Welsh Assembly Government Cartographics Unit produced a look up table for 2007 postcodes, where it was possible to calculate apportionment rates as above for each postcode. 2006 and 2005 look up tables were also supplied by the Welsh Assembly Government Cartographics Unit.

Best Fit to Postcode Boundaries

While the apportionment method has been used for most WIMD postcode level data, there were a small number of instances where out-of-date postcodes were contained in data records and the only option was to use the best fit method. A best fit method could be used as the ONS All Fields Postcode Directory (AFPD) holds historical postcodes. For future Indexes, the apportionment lookup will be built up over time to reduce the problem of out-of date postcodes.

Appendix B Population Denominators

Background

For the Welsh Index of Multiple Deprivation (WIMD) 2008, population estimates were needed for each of the 1,896 Lower layer Super Output Areas (LSOAs) to provide population denominators for many of the individual indicators. Estimates for the 'at risk' population were constructed using a combination of LSOA Small Area Population Estimates (SAPE) and LSOA level 2001 Census populations, both published by the Office for National Statistics (ONS).

The population estimates are on a usual residence basis and as such include household population and communal establishment population (e.g. students, persons in care establishments and children in local authority homes) but the prison population has been excluded. Prisoners are a special case in that they are not at risk for some forms of deprivation but are in any case isolated from the rest of society and as such do not contribute to the overall deprivation level.

The SAPE figures were first published by the ONS in April 2005. While work was under way within ONS to develop LSOA population estimates the timetable for this was well beyond the timetable for WIMD 2005, however, SAPE LSOA figures have been used in WIMD 2008.

Lower layer Super Output Areas

Lower layer SOAs (used for the WIMD 2008) contain a minimum population of 1,000 and a mean population of 1,500. These typically contain between 4 and 6 Census Output Areas, and are constrained by the boundaries of the Standard Table (ST) wards used for the 2001 Census. In Wales there are 1,896 LSOAs.

Other denominators

For some of the indicators, denominators were taken from the same datasets as the numerators. Indicators for which this applied were denominators for:

- The Key Stage 2, 3 and 4 indicators in the Education Domain.
- Absenteeism indicators for the Education Domain.
- No Qualifications indicator for the Education Domain.
- Limiting Long Term Illness from the Health Domain.
- Singleton Low Birth Weight from the Health Domain
- All indicators for the Housing Domain.
- All denominators for the Geographical Access to Services Domain were the numbers of address points within each LSOA.

Appendix C The Factor Analysis Technique

Factor Analysis Overview

Factor analysis is a method for assessing the extent to which a set of indicators may be measuring the same underlying construct or factor. The premise behind a one-common-factor model is that the underlying factor is imperfectly measured by each of the indicators in the dataset but that indicators that are most highly correlated with the underlying factor will also be highly correlated with each other. By analysing the correlation between indicators it is therefore possible to make inferences about the common factor and as a result estimate a 'factor score' for each LSOA. This score is derived from a set of weights for each of the indicators in the data set that is generated by the process of factor analysis. This factor score can then be used as the domain index.

Factor analysis has only been applied to four domains: Health, Education, Geographical Access to Services and Community Safety. Factor analysis is used in these domains because they contain indicators that measure, on potentially different metrics and with different levels of accuracy, a number of forms of that deprivation and therefore cannot otherwise easily be combined. The main reasons why Factor Analysis has been used are:

- Because the indicators are on different metrics and have different levels of accuracy, and so cannot simply be summed
- To ascertain the factor that underlies the indicators within the Domain
- To help take into account the problem of 'double counting' within a Domain

In the Employment and Income domains we can identify individuals who are or are not deprived in terms of the domain definition. The number of deprived people can then simply be summed and divided by a suitable denominator to create an area rate. This is not possible in the other six domains. These deprivations tend to present themselves in different ways at different times. Thus, for example, an individual is 'health deprived' if they die prematurely or are long-term sick. While the long-term sick may be more likely to die prematurely than others, these events do not occur to the same people at the same time. Typically such domains include data on people at different ages and stages e.g. in the education domain, lack of qualifications in the adult population as well as poor results at school level. Instead we hypothesise that there is an underlying factor at the local area level (e.g. health deprivation) that makes these different states likely to exist together in the same area. This underlying factor cannot be measured directly but can be identified through its effects on specific individual measures (e.g. premature death, long-term sickness, low birth-weight children etc.). We have therefore collected a number of indicators that measure, with different levels of accuracy, the effects of this underlying factor. By looking at the relationship between all these indicators the underlying factor can be identified and quantified.

Factor analysis also takes some account of the problem of 'double-counting' within domains. The Health, Education, Access to Services and Community Safety domains potentially contain indicators that overlap with each other. For example, in the Health domain, it is possible for an individual to have had cancer and also potentially to be included in the limiting long-term illness indicator. Combining data

using other methods such as 'z scores' more directly double-weights these cases by taking them all into account. Factor analysis, however, takes some account of this overlap in that an indicator may have a lower weight if the contribution it makes has already been taken into account.

The choice of maximum likelihood estimation method

WIMD 2008 follows the WIMD 2005 methodology and that applied by Oxford University for WIMD 2000 as well as the Indexes for the other three UK countries.

In Principal Components Analysis all variance in an indicator is analysed including measurement error (*error variance*) and the indicators' imperfect measurement of the underlying construct or constructs (*specific variance)*. This is because it does not attempt to separate *common variance* (i.e. variance shared between three or more indicators) from *unique variance* (i.e. specific variance and error variance). It assumes that an indicator is perfectly reliable and measured without error. It was therefore not appropriate to use the Principal Components method. The appropriate technique, where it is suspected that indicators are not perfectly reliable or measured without error, is *common factor analysis* of which Maximum Likelihood Factor (ML) analysis is a type.

Principal Factoring (PF) has, in the past, been the favoured method of common factor analysis but this was probably because of its relative computational simplicity. With the advent of high-powered computers more sophisticated methods, such as ML factor analysis, are now easily accomplished. PF has a number of disadvantages in comparison to ML factor analysis. The PF solution depends on the scale of measurement of the input indicators (i.e. depends on whether or not they have been standardised), which means that there is not one but an infinity of PF solutions, the choice among, which is arbitrary. The factor model itself is intrinsically scale free, and thus any procedures for its estimation should be scale invariant. ML is scale invariant. ML also treats the correlation matrix as a sample correlation matrix and attempts to explain variance in the *population* correlation matrix. This treatment of the data as a sampled dataset is consistent with the proposal, made throughout this project, that even 'census' indicators should be seen as a sample from a super-population.

Communality

This is the proportion of a variable's variance explained by a factor structure. A variable's commonality must be estimated prior to performing a factor analysis. A communality does not have to be estimated prior to performing a principal component analysis. Communality estimates are estimates of the proportion of common variance in a variable. *Prior communality estimates* are those which are estimated prior to the factor analysis. Common methods of prior communality estimation are to use (1) an independent reliability estimate, (2) the squared multiple correlation between each variable and the other variables, (3) the highest off-diagonal correlation for each variable, or (4) iterate by performing a sequence of factor analyses using the final communality estimates from one analysis as prior communality estimates for the next analysis. *Final communality estimates* are the sum

of squared loadings for a variable in an orthogonal factor matrix.
The default setting for communality prior estimates, Square Multiple Correlation, was used for WIMD 2008 calculations.

Calculation Process

The indicators were first transformed to the standard normal distribution. The transformed indicators were then entered into a 'one common factor Maximum Likelihood factor analysis'. Fuller's regression method was used to derive factor scores from the resulting solution. The process was undertaken in SAS Enterprise Guide and the following details the settings used.

- The normally transformed values for each of the domain indicators were entered as the analysis variables;
- Maximum likelihood factor analysis was chosen as the factoring method, for the reasons described above;
- The smallest eigenvalue was set to 1 because this is a commonly used indicator showing that sufficient factors have been extracted to reasonably explain the 'common variance' between the indicators.
- For prior communality estimates the method chosen was Squared Multiple Correlation with all other columns, as described above;
- For the rotation method, no rotation was selected as we are only looking for a single factor solution and rotation only applies if there two or more factors.

Appendix D Exponential Transformation of the Domain Indexes

The precise transformation involved is as follows. For any LSOA, denote its rank on the domain, scaled to the range [0,1], by R (with R=1/1896 for the least deprived, R=1896/1896=1 for the most deprived).

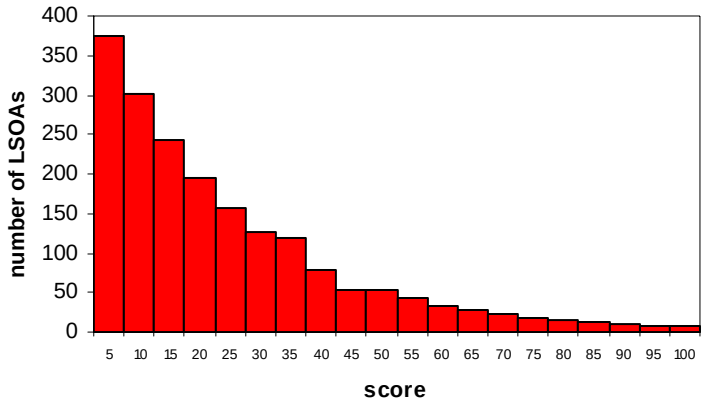
The transformed domain, (X) equals:

-23*log{1 - R*[1 - exp(-100/23)]}

where log denotes natural logarithm and exp the exponential or antilog transformation, and * denotes multiplication. This formula is straightforwardly calculated and is in fact simpler than the commonly-used transformation to a normal curve which necessitates the use of a look-up table. The resulting distribution is illustrated below in a histogram.

Each transformed domain has a range of 0 to 100, with a score of 100 for the most deprived LSOA. Ten percent of LSOAs have a score higher than 50. When transformed scores from different domains are combined by averaging them, the skewness of the distribution reduces the extent to which deprivation on one domain can be cancelled by lack of deprivation on another. For example, if the transformed scores on two domains are simply averaged, with equal weights, a (hypothetical) LSOA that scored 100 on one domain and 0 on the other would have a combined score of 50 and would thus be ranked at the 90th percentile. (Averaging the untransformed ranks, or after transformation to a normal distribution, would result in such a LSOA being ranked instead at the 50th percentile: the high deprivation in one domain would have been fully cancelled by the low deprivation in the other.) Thus the extent to which deprivation in some domains can be cancelled by lack of deprivation in others is, by design, reduced.

Histogram of a transformed domain



Appendix E List and details of indicators used

Income

Adults & Children in Income Support households
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.
Adults & Children in Pension Credit households
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.
Adults & Children in Income-Based Job Seekers Allowance households
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.
Adults & Children in Tax Credit (Child Tax Credit and Working Tax Credit) households below a low income threshold
Overall count by LSOA for August 2005 (2005-06 finalised awards) of in-work claimants of CTC and the out of work not claiming IS/JSA with equivalised income less than 60 per cent of the Wales median supplied by HMRC.
National Asylum Support Service (NASS) supported asylum seekers in Wales in receipt of subsistence only and accommodation support
Count by LSOA of asylum seekers in receipt of subsistence only support and in dispersed accommodation in Wales as at the end of June 2007 supplied by the Home Office.

Employment

Claimants of Unemployment-related benefits
Overall count of claimants by LSOA for the 12 moths from September 2006 to August 2007 obtained from the ONS (NOMIS).
Claimants of Incapacity Benefit / Severe Disablement Allowance
Overall count by LSOA for the four quarters from August 2006 to May 2007 supplied by DWP.
Participants on New Deal for Young People and Intensive Activity Period (for New Deal 25+) not included in unemployment-related benefit counts
Overall count by LSOA from 2005 supplied by DWP.
Participants on New Deal for Lone Parents
Overall count by LSOA from 2005 supplied by DWP.

Health

Limiting long-term illness (LLTI)
Data supplied by the ONS for all people usually resident in the area at the time of the 2001 Census who had a limiting long-term illness. A limiting long-term illness covers any long-term illness, health problem or disability that limits daily activities or work.
Standardised all-cause death rate
Counts of all-cause mortality by LSOA with 5-year age and sex breakdown for the 10 year period 1997-2006 supplied by the ONS. Data was age-sex standardised using direct standardisation and a 6-year average (2001 to 2006) of ONS population estimates.
Standardised cancer incidence rate
Count of all cases of cancer (all malignancies excluding non melanoma skin cancer) at LSOA level by sex, five year age band for the ten year period 1996-2005 supplied by Velindre NHS Trust.

Singleton low birth weights
Percentage of live singleton babies classed as having a low birth weight (defined as birthweight less than 2500g) using data for the 10 year period 1997 to 2006 supplied by the ONS.

Education

Key Stage 2, average point scores
Average point scores of pupils being taught in National Curriculum year group 6 in all schools (excluding Independent schools) in 2005, 2006 and 2007, measured using teacher assessments. Data is used from PLASC and postcodes matched to LSOAs using a look-up supplied by the Welsh Assembly Government Cartographics Unit.
Key Stage 3, average point scores
Average point scores of pupils being taught in National Curriculum year group 9 in all schools (excluding Independent schools) in 2005, 2006 and 2007, measured using teacher assessments. Data is used from PLASC and postcodes matched to LSOAs using a look-up supplied by the Welsh Assembly Government Cartographics Unit.
Key Stage 4, average point scores
Average point scores of pupils aged 15 as at 31 August 2005 and 2006 in all schools. All approved GCSE and GNVQ results are included for 2005 and 2006. Data is used from PLASC and postcodes matched to LSOAs using a look-up supplied by the Welsh Assembly Government Cartographics Unit.
Primary school absence rates
Average pupil proportions of total half-day sessions absence (including both authorised and unauthorised absence). Data on the number of school sessions missed due to authorised and unauthorised absence is collected from the start of the academic year up to the date of the late May bank holiday. The indicator used the total absences, for the 2005/06 and 2006/07 academic years, taken from PLASC.
Secondary school absence rates
Average pupil proportions of total half-day sessions absence (including both authorised and unauthorised absence). Data on the number of school sessions missed due to authorised and unauthorised absence is collected from the start of the academic year up to the date of the late May bank holiday. The indicator used the total absences, for the 2004/05, 2005/06 and 2006/07 academic years, taken from PLASC.
Proportion of people not entering Higher Education aged 18-19
The proportion of people not entering HE was then calculated using data on young participation rates for the 1998 to 2005 cohorts for Welsh LSOAs supplied by the Higher Education Funding Council for England (HEFCE).
Number of adults aged 25 – 59/64 with no qualifications
This indicator captures all adults aged 25 to retirement age (59 for women and 64 for men) who reported at the time of the 2001 Census as holding no qualifications or qualifications which do not reach the standard to be categorised as Level 1 qualifications.

Housing

Lack of central heating
The number of people living in households without central heating as a proportion of all people in households, derived from the 2001 Census.
Overcrowding (excluding all student households)
All people in households, excluding people living in all student households, living in households with an occupancy rating of -1 or less, as a proportion of all people in

households excluding people living in all student households, derived from the 2001 Census.

Physical Environment

Air Quality

Air pollution concentrations are low-resolution data estimated using models. They are estimated averages over an area. Concentration maps are validated against available monitoring data. The latest air quality maps were for 2006. Data supplied by the Environment Agency.

Air Emissions

Air pollution emissions are low-resolution data estimated using models and emissions inventories. They are estimated averages over an area. Concentration maps are validated against available monitoring data. The latest emission maps were for 2005. Data supplied by the Environment Agency.

Flood risk

Flood risk calculations and maps were based on the 2007 NAFRA database which includes flood risk, taking into account flood defences where these are known. Data supplied by the Environment Agency.

Proximity to waste disposal and industrial sites

Each site was assigned a proximity score which relates to the potential and actual deprivation the site could cause to the environment and the people living within it's vicinity. The OPRA score was used to assign the score for those sites where it is currently recorded. Where no OPRA score was available a proximity score was created based on best available information. Where no boundary data was available, the boundary of a site was assumed to be circular with a radius of 0.5km. Data supplied by the Environment Agency.

Geographical Access to Services

Food shops

The average time taken for every household in a LSOA to travel to a food shop by walking or using public bus services or both.

GP surgeries

The average time taken for every household in a LSOA to travel to a GP surgery by walking or using public bus services or both.

Primary schools

The average time taken for every household in a LSOA to travel to a primary school by walking or using public bus services or both.

Post office

The average time taken for every household in a LSOA to travel to a post office by walking or using public bus services or both.

Public library

The average time taken for every household in a LSOA to travel to a public library by walking or using public bus services or both.

Leisure centres

The average time taken for every household in a LSOA to travel to a leisure centre by walking or using public bus services or both.

NHS dentist

The average time taken for every household in a LSOA to travel to a NHS dentist by walking or using public bus services or both.

Secondary schools
The average time taken for every household in a LSOA to travel to a secondary school by walking or using public bus services or both.
Transport nodes
The average time taken for every household in a LSOA to travel to a transport node by walking or using public bus services or both.

Community Safety

Police Force Recorded Crime
Four crime types which affect individuals or businesses were selected; violence, burglary, criminal damage and theft. The incidents (crimes) were recorded by the police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent) and collected as counts by LSOA. The incidents were located to the point at which they occurred and allocated to the appropriate LSOA. Data are from 2005-2006 and 2006-2007.
Youth offenders
The number of offenders (aged 10 to 17 years of age) dealt with by the 18 Youth Offending Teams within Wales, using the person's address to allocate to a LSOA. Care homes and institutions were excluded from the counts. Data are from 2005-2006 and 2006-2007.
Adult offenders
The number of offenders (aged 18 and over) dealt with by the 4 Probation Service Areas (North Wales, Dyfed Powys, South Wales and Gwent) within Wales, using the person's address to allocate to a LSOA. Offenders residing in institutions were included in the count if a residential address was provided. Data are from 2005-2006 and 2006-2007.
Fire incidence
Incidents of primary fires with the addition of "derelict vehicle" fires were collected as counts by LSOA. Secondary "derelict vehicle" fires were only included if available at the LSOA level. Data for the Fire Incidence indicator is from 2005 and 2006.