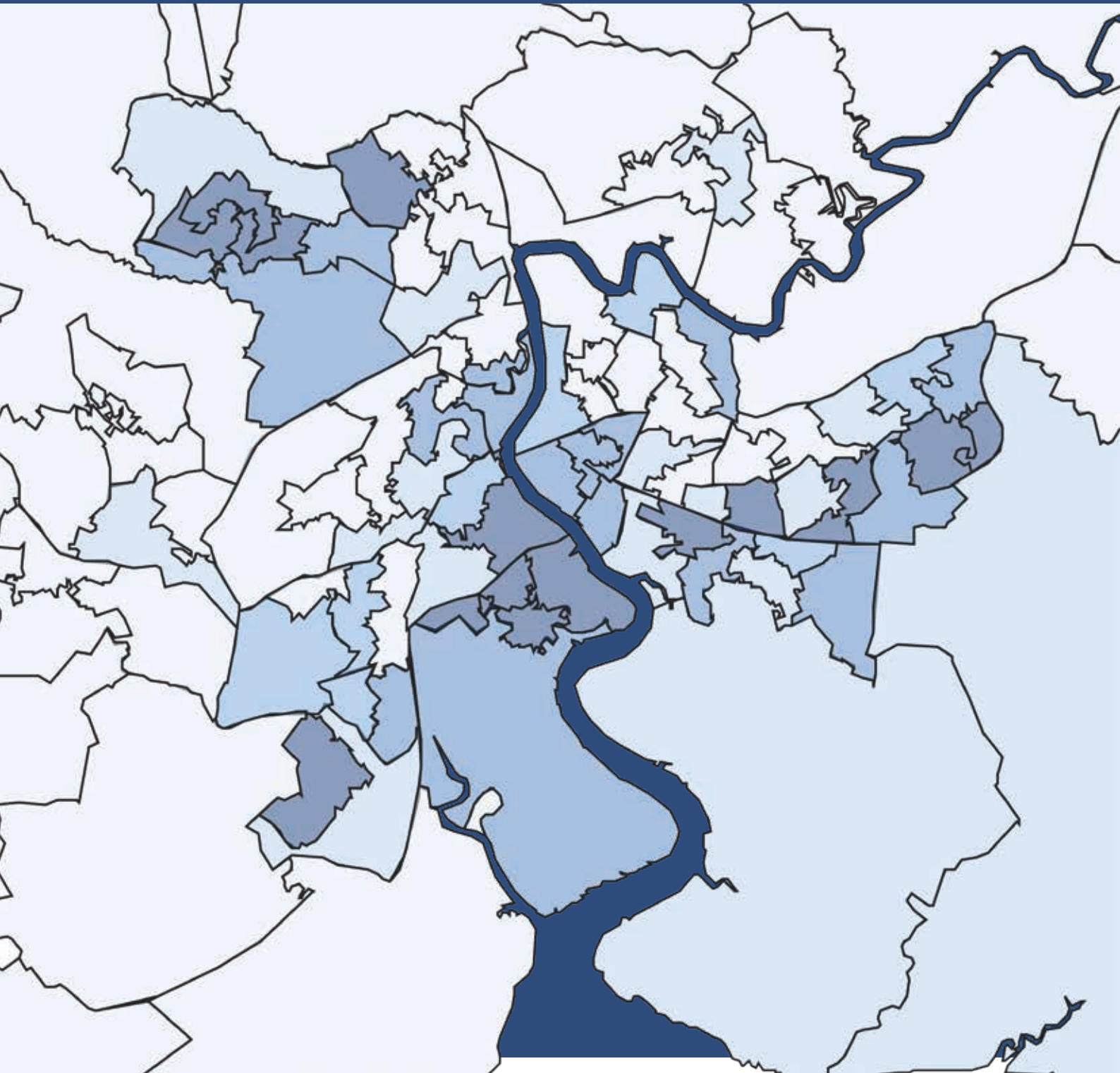
ystadegau ar gyfer cymru
statistics for wales

Llywodraeth Cymru
Welsh Government

Welsh Index of Multiple Deprivation (WIMD) 2019
Technical report



Contents

Contributions

We are grateful for the contributions of many people and organisations who have provided data and advice for WIMD 2019. We especially wish to thank the following people, who have helped to draft sections of this Technical Report:

- Access to Services and Physical Environment: Matthew Ricketts, Ordnance Survey
- Community Safety: David McLennan of Deprivation.org
- Health: Joe Hunt and Simon Scourfield, NHS Wales Informatics Service
- Housing: Rosie Winn and Fiona MacKenzie, Building Research Establishment
- Physical Environment: Natural Resources Wales

1. What is WIMD?

The Welsh Index of Multiple Deprivation (WIMD) is the Welsh Government's official measure of relative deprivation for small areas in Wales. It identifies areas with the highest concentrations of several different types of deprivation. WIMD ranks all small areas in Wales from 1 (most deprived) to 1,909 (least deprived). It is a National Statistic produced by statisticians at the Welsh Government.

WIMD is calculated from eight different domains (or types) of deprivation, each compiled from a range of different indicators. This Technical Report describes how WIMD 2019 was constructed and contains a full list of indicators and information about the indicators. Our WIMD Guidance document provides more information on the definition of deprivation, how to interpret and use WIMD. Our WIMD Results Report also provides examples of applications of WIMD.

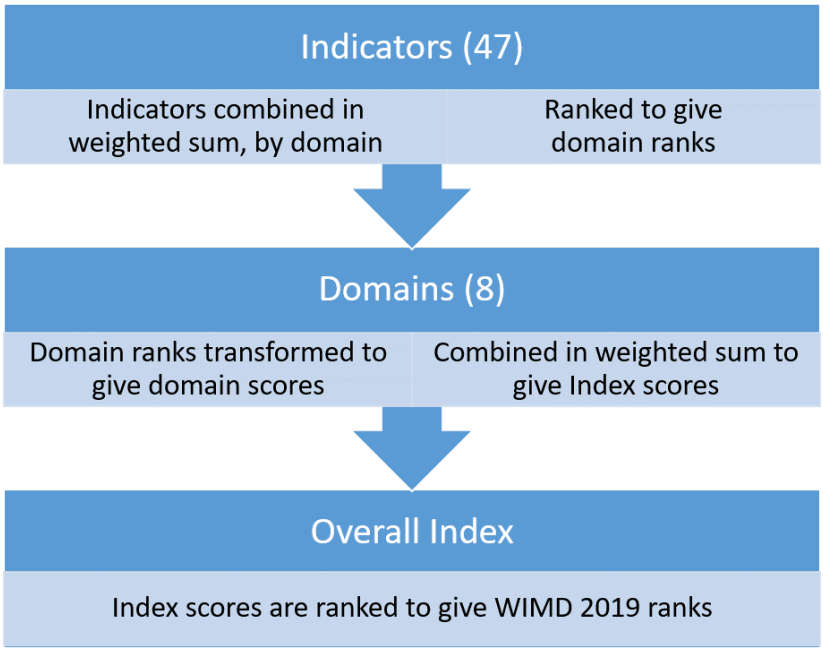
2. How the Index is Constructed

The Index has three main components:

- The 47 underlying indicator datasets
- Ranks for the eight separate domains (or types) of deprivation, created by combining relevant indicators within each domain
- Overall WIMD ranks, created by combining the domain ranks

All of these components are calculated for each of the small areas (LSOAs) in Wales. The indicator data, domain ranks, and overall ranks are published on StatsWales.

The diagram below provides an overview of how the different elements are combined together to produce the Overall Index. Further details follow beneath.



The methodology used within WIMD 2019 is broadly the same as for WIMD 2014, with the same eight domains or types of deprivation captured. However there have been some changes to the Index, including:

- changes to individual indicators, and the inclusion of new indicators within the Health, Education, Access to Services, Physical Environment and Housing domains
- small changes to the relative weights applied to the domains (or types) of deprivation.

Indicators

The domains are built up from sets of indicators. These are measurable quantities which capture the concept of deprivation for each domain (e.g. the percentage of working age people in receipt of employment-related benefits for the Employment domain, and a measure of educational attainment (KS4 average points score) in the Education domain). Indicators must be robust at the small area level and consistent across Wales. In practice, this means that the Index is based largely on administrative data, with a limited number of Census variables where appropriate administrative data are not available.

Weightings for indicators within domains are derived in several different ways, summarised as follows and explained in detail in each domain chapter:

- for Income and Employment domains, only one indicator exists therefore no weighting is required within the domain
- for the Housing domain, data for the two indicators are standardised to a normal distribution before being combined with equal weighting
- for the Physical Environment and Access to Services domains, indicators are grouped into two or more sub-domains, then the sub-domain ranks are exponentially transformed (see Annex A), before being combined in a weighted sum
- for the Education, Health, Access to Services (physical access sub-domain) and Community Safety domains, the statistical technique of factor analysis is used to calculate factor weights for the indicators (see Annex B)

Domain Scores

The overall Index and domain ranks are the main output for WIMD. As part of the process for calculating WIMD ranks, scores (domain and overall) are produced – see the WIMD Guidance for further advice on interpreting scores.

The individual domain ranks, calculated by ranking the weighted sum of domain indicators, are then exponentially transformed (see Annex A) to produce domain scores. Areas have scores (transformed ranks) ranging between 0 (least deprived) and 100 (most deprived) on each domain. The scores increase exponentially so that the most deprived areas have more prominence. This

reduces the extent to which deprivation in some domains can be cancelled by lack of deprivation in others.

The sets of domain scores are then weighted according to the respective domain weight and added together to produce the overall WIMD score, which is, in turn, ranked to provide the overall WIMD ranks.

Domain Weights

Domain weights reflect the importance of the domain as an aspect of deprivation, and the quality of the indicators available for that domain.

The domain weights for WIMD 2019 are shown below, alongside the weights used in 2014. Improvements to data used for the Housing and Health domains led to a decision to make small increases to their weights. To allow for this, weights for the Income and Employment domains were reduced slightly. However those domains retain the highest relative weights, since they are strong determinants of deprivation.

	WIMD 2019 domain weight	**WIMD 2014 domain weight**
Income	22%	23.5%
Employment	22%	23.5%
Health	15%	14%
Education	14%	14%
Access to Services	10%	10%
Housing	7%	5%
Community Safety	5%	5%
Physical Environment	5%	5%

3. WIMD Geographies

Super Output Areas

Following the 2001 Census of Population, the Office for National Statistics developed a geographic hierarchy called Super Output Areas (SOAs). They were designed to improve the reporting of small area statistics in England and Wales. The areas were reviewed, and some changes made, following the 2011 Census. Where possible, official statistics are published at the SOA geography.

There are three layers of SOA: Lower layer, Middle layer, and Upper layer. This is because disclosure requirements mean that some sets of data can be released for much smaller areas than others. To support a range of potential data requirements, it was decided to create these three SOA layers.

- A Lower layer SOA (known as an LSOA) must have a minimum population of 1,000. The mean size of all the Lower layer SOAs must be close to 1,600. They are built from groups of Census Output Areas (usually between four and six).
- A Middle layer SOA (MSOA) must have a minimum population of 5,000. The mean size of all the Middle layer SOAs must be close to 7,200.

Geographic unit for WIMD

The geographic areas used in the calculation of WIMD 2019 are the 1,909 Lower layer Super Output Areas (LSOAs) in Wales. LSOAs were used as the geographic unit in WIMD 2005, 2008, 2011 and 2014, and were designed for the reporting of small area statistics.

Although the overall WIMD rankings are only calculated at LSOA level, there are deprivation profiles for larger areas (like local authorities, local health boards and Assembly Constituency areas) available on StatsWales and the WIMD interactive tool. These look at the proportion of small areas within a larger area that are very deprived. Individual indicator data are also published at a range of geographies on StatsWales. Further information on individual indicator data can be found in the WIMD Indicator Guidance, which will be updated to reflect new indicators introduced in 2019 during early 2020.

For most domains, indicator data are allocated to an LSOA by the data suppliers as part of the collection process. However, data is provided at a lower geographical level for the travel time indicators in the Access to Services domain and for the Education and Community Safety domains. An explanation of how data were allocated to LSOAs for these domains is provided in Annex C.

4. Income domain

The purpose of this domain is to measure the proportion of people with income below a defined level. The income domain has a relative weight of 22% in the overall WIMD 2019 Index.

Domain construction

The income domain is made up of one indicator, containing four elements:

- Income-Related Benefit claimants
- Tax Credit recipients
- Supported Asylum Seekers
- People on Universal Credit (excluding those “working with no requirements”)

Dependent children of the above are also included. This indicator is expressed as a percentage of the residential population for each LSOA, with the prison population subtracted from the four LSOAs containing prisons in Wales (as at 2016).

The indicator values were ranked, then ranks were exponentially transformed to form domain scores for use in the calculation of the overall WIMD 2019.

Methodological changes between WIMD 2014 and WIMD 2019

The introduction of Universal Credit has had a significant impact on the measurement of the income domain of WIMD. Universal Credit is replacing most benefits and tax credits inputting to this domain. Once fully rolled out in all areas, those areas become directly comparable and can be ranked against one another even if eligibility conditions and levels of benefit differ from those relating to benefits currently within WIMD.

However, given the geographical nature and lengthy timescale of implementation, differences between Universal Credit and legacy benefits means we may not be comparing like with like at a given point during the rollout.

Although rollout of Universal Credit full service did not begin in Wales until April 2017, there had previously been implementation of Universal Credit for (mainly) single, new claimants of Job-Seekers Allowance.

To ensure consistency of data across areas, for WIMD 2019 we have used data at a point in time before the start of rollout of Universal Credit full service in Wales, which is the 2016-17 financial year. This does not impact much on the currency of the data compared to previous indices. Early, straightforward cases of Universal Credit claimants have been included.

Since WIMD 2014, other changes to the welfare system may mean that eligibility thresholds and criteria for some benefits have changed.

We will continue to work with the Department for Work and Pensions (DWP) on options for a suitable indicator using data on Universal Credit claimants in future indices.

Indicator

Publication	WIMD 2019
Domain	Income domain
Indicator Name	Percentage of people in income deprivation (in receipt of income-related benefits and tax credits)
Source	The Department for Work and Pensions (DWP): Income-Related Benefit claimants, and People on Universal Credit Her Majesty's Revenue and Customs (HMRC): Tax Credit recipients Home Office: Supported Asylum Seekers
Type of Indicator	Percentage
Denominator	Mid-2016 LSOA population estimates from the Small Area Population Estimates (SAPE) published by the Office for National Statistics (ONS), with the prison population as at mid-2016 (sourced from Ministry of Justice) subtracted for the four LSOAs with prisons in Wales.
Time period for WIMD 2019	Financial year 2016/17 for: 1) Income-Related Benefit claimants and their dependents 2) People on Universal Credit Supported Asylum Seekers as at 30th June 2017 2015/16 tax credits data as at 31st August 2015
Additional Notes	The income domain consists of a single composite indicator calculated from the following four elements: a) Income related benefit claimants; this includes Income Support claimants, Jobseeker's Allowance claimants, Pension Credit claimants, and Income Based Employment and Support Allowance claimants; and their dependent children. Individuals who are in receipt of multiple benefits are only counted once within this element. b) The number of children and adults within families that are in receipt of Working Tax Credits and Child Tax Credits with an income which is less than 60% of the median income for Wales (Before Housing Costs). c) Supported Asylum Seekers means those who are supported under Section 95, and their dependent children. d) People on Universal Credit started to be included from the 2016 indicator data update, and their dependent children were included from 2017. We

	exclude those who are “working with no requirements” according to the dataset’s “conditionality” marker. The different elements are summed and expressed as a percentage of the total residential population for the LSOA (with the prisons population subtracted from the four LSOAs with prisons in Wales).
Comparability with WIMD 2014	Since WIMD 2014, changes to the welfare system may mean that eligibility thresholds and criteria for some benefits have changed. Also see earlier description of methodological changes in terms of the introduction of Universal Credit. The impact of this change is likely to be minimal during the reference period used for WIMD 2019 data.

5. Employment domain

The purpose of this domain is to capture lack of employment. This covers involuntary exclusion of the working-age population from work, including those who cannot work due to ill-health or who are unemployed, but actively seeking work. The employment domain has a relative weight of 22% in the overall WIMD 2019 Index.

Domain construction

The employment domain is made up of one indicator, containing four elements:

- Jobseeker's Allowance (JSA);
- Employment and Support Allowance (ESA);
- Incapacity Benefit (replaced Severe Disablement Allowance)
- Universal Credit (UC) and not in employment

This indicator is expressed as a percentage of the residential working-age population for each LSOA, with the prison population subtracted from the four LSOAs containing prisons in Wales (as at 2016).

Methodological changes between WIMD 2014 and WIMD 2019

The introduction of Universal Credit has had a significant impact on the measurement of the employment domain of WIMD. Universal Credit is replacing most benefits and tax credits inputting to this domain. Once fully rolled out in all areas, those areas become directly comparable and can be ranked against one another even if eligibility conditions and levels of benefit differ from those relating to benefits currently within WIMD.

However, given the geographical nature and lengthy timescale of implementation, differences between Universal Credit and legacy benefits means we may not be comparing like with like at a given point during the rollout.

Although rollout of Universal Credit full service did not begin in Wales until April 2017, there had previously been implementation of Universal Credit for (mainly) single, new claimants of Job-Seekers Allowance.

To ensure consistency of data across areas, for WIMD 2019 we have used data at a point in time before the start of rollout of Universal Credit full service in Wales, which is the 2016-17 financial year. This does not impact much on the currency of the data compared to previous indices. Early, straightforward cases of Universal Credit claimants have been included.

Since WIMD 2014, other changes to the welfare system may mean that eligibility thresholds and criteria for some benefits have changed.

We will continue to work with the Department for Work and Pensions (DWP) on options for a suitable indicator using data on Universal Credit claimants in future indices.

Indicator

Publication	WIMD 2019
Domain	Employment domain
Indicator Name	Percentage of working-age population in employment deprivation (in receipt of Employment-related benefits)
Source	The Department for Work and Pensions (DWP)
Type of Indicator	Percentage
Denominator	Working-age population for the LSOA from the Mid-2016 Small Area Population Estimates published by the Office for National Statistics (ONS), with the prison population as at mid-2016 (sourced from Ministry of Justice) subtracted for the four LSOAs with prisons in Wales.
Time period for WIMD 2019	Financial Year 2016/17 for all four elements of the indicator
Additional Notes	This is calculated from a count of individuals (i.e. those who claim multiple benefits are only counted once) entitled to: - Jobseeker's Allowance (JSA); - Employment and Support Allowance (ESA); - Incapacity benefit (replaced Severe Disablement Allowance); - Universal Credit (UC) and not in employment This indicator is expressed as a percentage of the working-age population for the LSOA (with the prisons population subtracted from the four LSOAs with prisons in Wales).
Comparability with WIMD 2014	Since WIMD 2014, changes to the welfare system may mean that eligibility thresholds and criteria for some benefits have changed. Also see earlier description of methodological changes in terms of the introduction of Universal Credit. The impact of this change is likely to be minimal during the reference period used for WIMD 2019 data.

6. Health domain

The purpose of this domain is to capture deprivation relating to the lack of good health.

The domain has a relative weight of 15% in the overall index. This has increased from 14% in the 2014 index, due to the addition of three new indicators capturing a broader range of health-related conditions.

Domain construction

There are seven indicators in the health domain, weighted as follows.

- 31% GP-recorded chronic conditions (rate per 100), age-sex standardised
- 30% Limiting long-term illness (rate per 100), age-sex standardised
- 18% Premature deaths (rate per 100,000), age-sex standardised
- 10% GP-recorded mental health conditions (rate per 100), age-sex standardised
- 4% Cancer incidence (rate per 100,000), age-sex standardised
- 4% Low birth weight (live single births less than 2.5 Kg, percentage)
- 3% Children aged 4-5 who are obese (percentage)

An indirect age-sex standardisation process was applied to five of the indicators, in order to adjust for different age and sex distributions amongst small area populations. For example, one might expect to observe a higher rate of deaths in an aging population than in one consisting predominantly of young families. Standardisation attempts to adjust for these differences in population (see Annex D).

For every indicator, each area was ranked in order, with the most deprived area ranked 1 and the least deprived area ranked 1,909. These ranks were assigned to a normal distribution, with low ranks receiving a low normalised value. Factor analysis was then used to calculate the indicator weights. As with all domains, the final domain ranks were exponentially transformed, to form domain scores for use in the calculation of the overall WIMD 2019.

Methodological changes

There have been considerable developments in the health domain since WIMD 2014 with the health status of communities now being captured through a wider range of data sources. To help illustrate this, we have grouped our indicators into the three categories below:

Indicators that relate to children

As well as updating the measure on low birth weight, we have also included a new measure on children aged 4 and 5 who are obese according to the Child Measurement Programme. We chose

to include this measure given the links between being an unhealthy weight in childhood and long term health problems in adulthood.

Indicators that capture specific diagnosed conditions

In WIMD 2019, we have expanded the range of indicators from cancer incidence alone, to include measures on the prevalence of GP-recorded chronic conditions and mental health conditions.

Indicators that capture the status of the population's health more broadly

The indicator on limiting long term illness has not changed since 2014, since it is census-based.

The all-cause death rate, was included in WIMD 2011 and WIMD 2014 as a proxy for morbidity. We are now including an indicator on the all-cause rate of deaths for those that are under the age of 75, the premature death rate, instead of the previous indicator. The premature mortality measure is more sensitive to poor health that is manifest through lower life expectancy and so is more relevant to health deprivation in a population.

Indicators

Publication	WIMD 2019
Domain	Health
Indicator Name	People with a GP-recorded diagnosis of a chronic condition
Source	NHS Wales Informatics Service
Type of Indicator	Rate per 100, indirectly age-sex standardised
Denominator	Small Area Population Estimates, 2018, Office for National Statistics (ONS), minus prisoner numbers 2018, Ministry of Justice via ONS
Time period for WIMD 2019	31 March 2019
Additional Notes	The numerator was based on counts of people with diagnoses for conditions from a defined list of disease registers and sub-indicators obtained from GP practices in Wales. It measures the number of people with a current diagnosis of one or more of the conditions listed below: • Coronary Heart Disease • Chronic Obstructive Pulmonary Disease • Stroke and Transient Ischaemic Attack • Peripheral Arterial Disease • Chronic Kidney Disease • Diabetes Mellitus (type 1 for all ages, type 2 and other types for people aged 17+) • Epilepsy These counts were de-duplicated so that patients with more than one condition were not counted twice. Patient level data was aggregated to small areas (LSOAs), according to patient addresses, so that prevalence is based on where people live rather than where they are registered with a GP. See Annex E for more detail on this new indicator.
Comparability with WIMD 2014	New indicator

Publication	WIMD 2019
Domain	Health
Indicator Name	Limiting Long-Term Illness
Source	2011 Census, Office for National Statistics
Type of Indicator	Rate per 100, indirectly age-sex standardised
Denominator	2011 Census population estimates
Time period for WIMD 2019	2011
Additional Notes	This indicator covers any long-term illness, health problem or disability that limits daily activities or work, and includes all usual residents.
Comparability with WIMD 2014	The same data was used, since there was no Census update (or alternative source at small area level) in-between indices.

Publication	WIMD 2019
Domain	Health
Indicator Name	Premature Death Rate (deaths of those under the age of 75)
Source	Death registrations, Office for National Statistics
Type of Indicator	Rate per 100,000, indirectly age-sex standardised
Denominator	Small Area Population Estimates, 2008-2017 average, Office for National Statistics
Time period for WIMD 2019	2009-2018 average
Additional Notes	Poor health can manifest itself in lower life expectancy, which can be captured through age and sex standardised death rates.
Comparability with WIMD 2014	Previously, WIMD used an all age (and all cause) death rate. Therefore figures are not comparable.

Publication	WIMD 2019
Domain	Health
Indicator Name	People with a GP-recorded diagnosis of a mental health condition
Source	NHS Wales Informatics Service
Type of Indicator	Rate per 100, indirectly age-sex standardised
Denominator	Small Area Population Estimates, 2018, Office for National Statistics (ONS), minus prisoner numbers 2018, Ministry of Justice via ONS
Time period for WIMD 2019	31 March 2019
Additional Notes	The numerator was based on counts of people with diagnoses from a defined list of disease registers and sub-indicators obtained from GP practices in Wales. It measures the number of people with a current diagnosis of one or more of the conditions listed below: • Depression • Low mood (patients with record of low mood and an active repeat prescription for an anti-depressant) • Anxiety disorder (including panic disorders) • Dementia • Severe mental illnesses (schizophrenia, bipolar affective disorder and other psychoses) These counts were de-duplicated so that patients with more than one condition were not counted twice. Patient level data was aggregated to small areas (LSOAs), according to patient addresses, so that prevalence is based on where people live rather than where they are registered with a GP. See Annex E for more detail on this new indicator.
Comparability with WIMD 2014	New indicator

Publication	WIMD 2019
Domain	Health
Indicator Name	Cancer Incidence
Source	Welsh Cancer Intelligence and Surveillance Unit, Velindre NHS Trust
Type of Indicator	Rate per 100,000, indirectly age-sex standardised
Denominator	Small Area Population Estimates, 2007-2016 average, Office for National Statistics
Time period for WIMD 2019	2007-2016 average
Additional Notes	Count of all cases of cancer includes all malignancies, excluding non-melanoma skin cancer.
Comparability with WIMD 2014	Not comparable. The technique of indirect standardisation involves using different population structures to calculate each figure. It is not legitimate to make comparisons over time.

Publication	WIMD 2019
Domain	Health
Indicator Name	Low birth weight
Source	Birth registrations, Office for National Statistics (ONS)
Type of Indicator	Percentage
Denominator	Number of singleton live births, ONS
Time period for WIMD 2019	2009-2018 average
Additional Notes	The indicator on Low Birth Weight measures the percentage of live single births less than 2.5 kg (5.5lb), which is classified as a low birth weight. This can be linked to the mother's lifestyle and health as well as cause problems for the child in later life.
Comparability with WIMD 2014	Comparable

Publication	WIMD 2019
Domain	Health
Indicator Name	Children aged 4-5 who are obese
Source	Child Measurement Programme, Public Health Wales
Type of Indicator	Percentage
Denominator	Number of children aged 4-5, Child Measurement Programme
Time period for WIMD 2019	2012-13 to 2017-18 average
Additional Notes	The indicator on children who are obese measures the proportion of reception aged children (those aged 4 and 5) who are obese. This data is provided by Public Health Wales (PHW), and captured through the Child Measurement Programme (CMP) for Wales. The smallest level that PHW publish the CMP data at is Middle layer Super Output Area (MSOA) level, rather than LSOA. This is because of concerns around the possible misuse of data to identify specific areas where the highest percentage of obese children live and the possibility of identifying individuals when drilling down to small numbers. For these reasons, whilst LSOA level rates are used in the domain and Index calculations, we only publish indicator values for MSOAs and higher level geographies as part of WIMD indicator datasets. Pupil's home addresses were used to identify the LSOA in which children live rather than the LSOA of their school. Obesity is calculated using the age and sex-specific body mass index (BMI) centiles (which includes height information) calculated using the British 1990 growth reference (UK90) (from a method proposed by Cole et al (1995)). Children who fall in the 95th centile or above are considered to be obese.
Comparability with WIMD 2014	New indicator

7. Education domain

The purpose of this domain is to capture the extent of deprivation relating to education, training and skills. It is designed to reflect educational disadvantage within an area in terms of lack of qualifications or skills. The domain has a relative weight of 14% in the overall index.

Domain construction

There are six indicators in the education domain, weighted as follows.

- 9.9% Foundation Phase Average Point Score
- 11.7% Key Stage 2 Average Point Score
- 27.6% Key Stage 4 Average Point Score for Core Subjects
- 21.7% Repeat Absenteeism
- 13.2% Proportion of Key Stage 4 leavers entering Higher Education
- 16.0% Number of Adults aged 25-64 with No Qualifications

The weight of the number of adults aged 25-64 with no qualifications indicator was capped at 16%. Factor analysis was then used to allocate the weights of the remaining indicators.

For every indicator, each LSOA was ranked in order, with the most deprived LSOA ranked 1 and the least deprived LSOA ranked 1,909. These ranks were assigned to a normal distribution, with low ranks receiving a low normalised score. They were then combined using the weights above. As with all domains, the final domain ranks were exponentially transformed to form domain scores (see Annex A).

Methodological changes

There have been several methodological changes to the education domain between WIMD 2014 and WIMD 2019. A full list of the changes is outlined below.

Three of the six indicators in the domain were present in WIMD 2014:

- Key Stage 2 Average Point Score
- Repeat Absenteeism
- Number of Adults aged 25-64 with No Qualifications

A Foundation Phase indicator based upon the results of teacher assessments for pupils taught in National Curriculum Year Group 2 has been included in WIMD 2019. It is intended to complement the Key Stage 2 indicator to measure education deprivation at primary school level.

The two Key Stage 4 indicators included in WIMD 2014 have been replaced with a single indicator. This indicator has been constructed using a new methodology and is, therefore, not comparable with the previous indicators used in WIMD 2014.

The Proportion of People aged 18-19 not entering Higher Education indicator has been replaced with an indicator measuring the Proportion of Key Stage 4 Leavers entering Higher Education. Whilst these indicator measure similar things, they are not directly comparable.

Indicators

<table>
<tr><td>Publication</td><td>WIMD 2019</td></tr>
<tr><td>Domain</td><td>Education</td></tr>
<tr><td>Indicator Name</td><td>Foundation Phase Average Point Score</td></tr>
<tr><td>Source</td><td>Pupil Level Annual School Census (PLASC) and National Data Collection (NDC)</td></tr>
<tr><td>Type of Indicator</td><td>Points score</td></tr>
<tr><td>Denominator</td><td>Total number of pupils in National Curriculum Year Group 2</td></tr>
<tr><td>Time period for WIMD 2019</td><td>Three year average for Academic Years 2015/16, 2016/17 and 2017/18</td></tr>
<tr><td>Additional Notes</td><td>Average point scores for pupils are assessed by teachers in Year 2 (final year of the Foundation Phase) for each of the Foundation Phase Areas of Learning (AoL). Postcode data from PLASC is matched to LSOAs, using a postcode to LSOA look-up. A 3-year average is used to reduce the impact of having small numbers of pupils at LSOA level. As not all children are assessed in Welsh as a first language at Foundation Phase, the highest score in ‘Language, literacy and communication skills – English’ and ‘Language, literacy and communication skills – Welsh’ was taken along with the scores in ‘Mathematical Development’ and ‘Personal and social development, well-being and cultural diversity’, to provide comparability across Wales. The indicator only includes maintained schools.

The number of points received for each award is in the following table.
<table>
<tr><th>Foundation Phase test outcome</th><th>Points: all AoL</th></tr>
<tr><td>Disapplied[1]</td><td>0</td></tr>
<tr><td>Absent[2]</td><td>0</td></tr>
<tr><td>Not awarded</td><td>0</td></tr>
<tr><td>Working towards FP outcomes</td><td>3</td></tr>
<tr><td>Foundation Phase Outcome Bronze</td><td>3</td></tr>
<tr><td>Foundation Phase Outcome Silver</td><td>3</td></tr>
<tr><td>Foundation Phase Outcome Gold</td><td>3</td></tr>
<tr><td>Foundation Phase Outcome 1</td><td>9</td></tr>
</table>
</td></tr>
</table>

	Foundation Phase Outcome 2 15 Foundation Phase Outcome 3 21 Foundation Phase Outcome 4 27 Foundation Phase Outcome 5 33 Foundation Phase Outcome 6 39 Performance Above Foundation Phase Outcome 6 45 **Notes on the points scoring table.** 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.
Comparability with WIMD 2014	New indicator

Publication	WIMD 2019
Domain	Education
Indicator Name	Key Stage 2 Average Point Score
Source	Pupil Level Annual School Census (PLASC) and National Data Collection (NDC)
Type of Indicator	Points Score
Denominator	Total number of pupils in National Curriculum Year Group 6
Time period for WIMD 2019	Three year average for Academic Years 2015/16, 2016/17 and 2017/18

Additional Notes

Average point scores for pupils are assessed by teachers in Year 6 (final year of primary school). Postcode data from PLASC is matched to LSOAs, using a postcode to LSOA look-up. A 3-year average is used to reduce the impact of having small numbers of pupils at LSOA level. 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. The indicator only includes maintained schools.

In WIMD 2014 this indicator included data on Welsh-domiciled year 7 pupils attending schools in England. However, due to divergence between the KS2 curriculum in England and Wales, the WIMD 2019 indicator does not include such pupils.

The number of points awarded for each outcome level is detailed in the following table:

KS2 test outcome (Level)	Points: all subjects
Disapplied[1]	0
Absent[2]	0
Level 'N' (no information available)	0
National Curriculum Outcome 1	3
National Curriculum Outcome 2	3
National Curriculum Outcome 3	3
National Curriculum Level 1	9
National Curriculum Level 2	15
National Curriculum Level 3	21
National Curriculum Level 4	27
National Curriculum Level 5	33
National Curriculum Level 6	39
National Curriculum Level 7	45
National Curriculum Level 8	51
Exceptional Performance	57

Notes on the points scoring table.

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.

Comparability with WIMD 2014	Comparable

Publication	WIMD 2019
Domain	Education
Indicator Name	Key Stage 4 Average Point Score for Core Subjects
Source	National Pupil Database (NPD), Pupil Level Annual School Census (PLASC) and Welsh Examinations Database (WED)
Type of Indicator	Points Score
Denominator	Total number of pupils in National Curriculum Year Group 11
Time period for WIMD 2019	Three year average for Academic Years 2015/16, 2016/17 and 2017/18

<table>
<tr><td>Additional Notes</td><td>This is a new indicator and replaces the two Key Stage 4 indicators included in WIMD 2014. It measures the average point score of Year 11 pupils based on the grades achieved in the core subjects of GCSE English or Welsh, GCSE Mathematics and GCSE Science. Postcode data from PLASC is matched to LSOAs, using a postcode to LSOA look-up. A 3-year average is used to reduce the impact of having small numbers of pupils at LSOA level

The points score awarded for each grade is detailed in the following table:
<table>
<tr><th>GCSE grade</th><th>Points</th></tr>
<tr><td>U</td><td>0</td></tr>
<tr><td>G</td><td>16</td></tr>
<tr><td>F</td><td>22</td></tr>
<tr><td>E</td><td>28</td></tr>
<tr><td>D</td><td>34</td></tr>
<tr><td>C</td><td>40</td></tr>
<tr><td>B</td><td>46</td></tr>
<tr><td>A</td><td>52</td></tr>
<tr><td>A*</td><td>58</td></tr>
</table>

Results data for pupils domiciled in Wales but attending a school in England has been provided by the Department for Education (DfE) and included in the calculation of this indicator. The following table details the points score awarded for results using the new 9-1 grading system:
<table>
<tr><th>GCSE grade</th><th>Points</th></tr>
<tr><td>1</td><td>16</td></tr>
<tr><td>2</td><td>24</td></tr>
<tr><td>3</td><td>32</td></tr>
<tr><td>4</td><td>40</td></tr>
<tr><td>5</td><td>44</td></tr>
<tr><td>6</td><td>48</td></tr>
<tr><td>7</td><td>52</td></tr>
<tr><td>8</td><td>55</td></tr>
<tr><td>9</td><td>58</td></tr>
</table></td></tr>
<tr><td>Comparability with WIMD 2014</td><td>New indicator</td></tr>
</table>

Publication	WIMD 2019
Domain	Education
Indicator Name	Repeat Absenteeism
Source	National Pupil Database (NPD) and Pupil Level Annual School Census (PLASC)
Type of Indicator	Percentage
Denominator	Total numbers of primary and secondary school pupils
Time period for WIMD 2019	Academic Years 2015/16, 2016/17 and 2017/18.
Additional Notes	This indicator measures the percentage of primary and secondary school pupils missing more than 15% of school sessions and is used as a proxy for persistent absenteeism. Postcode data from PLASC is matched to LSOAs, using a postcode to LSOA look-up. Data are based on all pupils of statutory school age attending a maintained school. Data on pupils domiciled in Wales but attending a school in England has been provided by the Department for Education (DfE) and included in the calculation of this indicator.
Comparability with WIMD 2014	Comparable

Publication	WIMD 2019
Domain	Education
Indicator Name	Proportion of Key Stage 4 leavers entering Higher Education
Source	Pupil Level Annual Schools Census (PLASC), Higher Education Statistics Authority (HESA) Record, Lifelong Learning Wales Record (LLWR).
Type of Indicator	Percentage
Denominator	Total number of pupils in National Curriculum Year Group 11
Time period for WIMD 2019	4 year average for Academic Years 2011/12 to 2014/15.
Additional Notes	This is a new indicator and replaces the Proportion of People Aged 18-19 not entering Higher Education indicator included in WIMD 2014. Pupil-level data from PLASC has been matched to HESA and LLWR data to measure the proportion of KS4 pupils who, at some point in the subsequent three years after leaving Year 11, entered Higher Education. Higher Education provision is defined for the purposes of this indicator as any programme of learning above level 3 – i.e. it includes undergraduate degrees, higher apprenticeships
Comparability with WIMD 2014	New indicator

Publication	WIMD 2019
Domain	Education
Indicator Name	Number of adults aged 25-64 with No Qualifications
Source	2011 Census, Office for National Statistics (ONS)
Type of Indicator	Percentage
Denominator	Total population aged 25-64
Time period for WIMD 2019	2011
Additional Notes	This indicator was included in WIMD 2014 and is the only measure of educational deprivation amongst the adult population. Due to recent changes to the state pension age, this indicator looks at those aged 25-64.
Comparability with WIMD 2014	Comparable

8. Access to Services domain

The purpose of this domain is to capture deprivation as a result of a household's inability to access a range of services considered necessary for day-to-day living, both physically and online. This covers both material deprivation (e.g. not being able to get food) and social aspects of deprivation (e.g. not being able to attend afterschool activities). Poor access to services is a factor that compounds other types of deprivation that exist within an area.

The access to services domain has a weighting of 10% in WIMD 2019.

Domain construction

The access to services domain measures physical access through travel times to 8 services using public transport and 9 services using private transport. Public transport includes travel by: public bus, public train, foot and national coach. Private transport is considered to be transport by private car. An access to digital services indicator has also been added in WIMD 2019, measuring access to superfast broadband.

The travel time indicators are a weighted average of the private and public transport times to each service (with the exception of petrol stations). The weights are calculated for each LSOA using data from the 2011 Census on car ownership and the number of adults aged 17 and over.

The digital access indicator is calculated as the proportion of homes and small businesses unable to receive fixed line broadband at a download sync speed of 30Mb/s or higher (superfast broadband). This includes properties with no broadband coverage.

For every indicator, each LSOA was ranked in order, with the most deprived LSOA ranked 1 and the lowest deprived LSOA ranked 1,909. The physical and digital aspects are treated as two sub-domains, with the 9 travel time indicators weighted together using factor analysis (see resultant weights below). Data for the two sub-domains was exponentially transformed before being combined, with the digital services sub-domain given a weight of 10%. As with all domains, the final domain ranks were then exponentially transformed, to form domain scores for use in the calculation of the overall WIMD 2019.

- 21.5% Pharmacy
- 16.9% Food shop
- 14.2% General Practitioner (GP)
- 10.2% Post office
- 10.0% % Unavailability of broadband at 30Mb/s
- 7.5% Primary school
- 7.3% Public library
- 5.3% Sports Facility
- 4.7% Secondary school
- 2.4% Petrol station (private transport only)

Methodological changes

As well as introducing a new indicator on access to digital services, we have updated the WIMD 2014 suite of indicators measuring travel times to key services. Whilst the principles of the methodology for measuring travel times have not changed between WIMD 2014 and WIMD 2019 – the technical toolkits used have undergone rigorous change to seek to increase the accuracy, robustness and repeatability of the data. The changes for WIMD 2019 are as follows:

Improvements to granularity of private routing

The algorithm developed for the Private Transport aspect of the domain has been improved by conducting each route traversal from each of the services to the doorstep of every residential address. To achieve this the road network geometry was split at the point on the road network that is closest to each residential address, rather than building a network graph based on the standard link-node structure of the dataset.

Improvements to granularity of public transport routing

The Public Transport methodology has changed from a gridded catchment approach to that of an isochrone approach. In 2014 a coarse grid (500m resolution) was cast over Wales to which public transport travel times were attributed from each service. For 2019, an isochrone approach has been employed. Achieving a finer level of detail, isochrones (shapes representing how far one can travel in a given time) are produced from 0 to 90 minutes in 5-minute intervals to accurately capture the areas of Wales that can and cannot be accessed via. the Public Transport network.

Improvements to the definition of residential dwelling to be as encompassing to the population of Wales as possible

For the 2019 definition of a residential address the domestic council tax record is utilised to capture all forms of residential dwellings across Wales that may or may not be classified as residential as their primary function. This particularly benefits residential dwellings that exist within farm estates, where the address may be categorised as a commercial farm but within also contains a residential dwelling where the citizen is still in receipt of services from the local authority.

Types of services modelled remain unchanged from WIMD 2014, and as such, data sources, the number of sites and the definitions adhered to are mostly comparable. However, the Food shops indicator has been enhanced to include certain chains of Frozen Food retailers who adhere to our basic provisions definition. Leisure Centres have been renamed to Sports Facilities to encompass the wider service offering now modelled.

Due to the above, care should be taken in interpreting changes in the travel time indicator values, which are not strictly comparable to older data. The main effect of the update is to raise the lower travel times, leading to notable movements between relative ranks at the least deprived end. More generally, reasons for changes include:

- changes to the methodology used to calculate the travel times by both private and public transport (see above)
- the opening and closure of services since 2014
- changes in public transport and road networks since 2014
- changes to location of residential dwellings in an area (e.g. new housing) since 2014

Indicators

Physical Access

Public transport travel times (walking and using a public bus, public train or national coach) to the nearest access point for a given service were calculated using propeR.

Private transport travel times to the nearest access point for a given service were calculated using the pgRouting library within PostGIS. The vehicular network was captured in the form of Ordnance Survey MasterMap Highways Network, with the average vehicular speed based on Average Speed Data from Basemap.

Publication	WIMD 2019
Indicator Name	Physical access – travel times to key services
Domain	Access to Services domain
Source	Welsh Government
Type of Indicator	Travel time (minutes), by public and private transport, to a given service
Denominator	N/A
Time period for WIMD 2019	Average speed data reflects the average vehicular speed across 2018. The road geometry, public transport timetables and all service origins were obtained in March 2019.
Additional Notes	Pharmacy - a store where medicinal drugs are dispensed and sold, this includes pharmacies within a larger complex or supermarket. Food shop - a store where you can purchase bread and milk. This includes convenience stores, independent supermarkets, frozen food retailers and supermarkets.

	General Practitioner (GP) surgery - facilities where NHS GPs are registered to practice. Public library - libraries that are open to the public. Mobile libraries are not included, due to a lack of geographical data. Post office - includes all static post offices. Primary school - schools that deliver education to 5-11 year olds. Travel times included in the calculation were limited to those where a child within the postcode was enrolled at a primary school. School enrolment data were sourced from PLASC. Sports Facility - a non-private and/or non-commercial (i.e. free or pay-for-play) site containing either one of the following: Sports Hall, Studio, Grass Pitch, Synthetic Turf Pitch, Swimming Pool & Health and Fitness Suite , or two or more of the following (can be in conjunction with one of previous): Squash Court, Outdoor Tennis Court, Outdoor Bowling Green, Indoor Tennis Centre, Indoor Bowls Petrol station - only included in the private transport element of the indicator. Secondary school - schools that deliver education to 11-16 year olds. Travel times included in the calculation were limited to those where a child within the postcode was enrolled at a secondary school. School enrolment data were sourced from PLASC. For detailed information of the parameters used to calculate physical access to services, see Annex F.
Comparability with WIMD 2014	Types of services modelled remain unchanged. Data sources, number of sites and definitions adhered to are mostly comparable. Food retailers have been enhanced to include certain chains of Frozen Food retailers who adhere to our basic provisions definition. Leisure Centres have been renamed to Sports Facilities to encompass the wider service offering now modelled.

Digital Access

Publication	WIMD 2019
Indicator Name	% Unavailability of broadband at 30Mb/s
Domain	Access to Services domain
Source	Ofcom Connected Nations Report, Spring 2019 (analysis if operator data)
Type of Indicator	Proportion of homes and small businesses unable to receive fixed line broadband at a download sync speed of 30Mb/s or higher (superfast broadband).
Denominator	Number of premises (homes and small businesses), Ordnance Survey (OS) Addressbase® Premium dataset (August 2018, Epoch 60).
Time period for WIMD 2019	As at January 2019
Additional Notes	This includes properties with no broadband coverage. For further information on the Ofcom data underlying the digital access indicator, see Annex G. Note that the Ofcom “Connected Nations” report is concerned with residential broadband services, which may also be used by small businesses. It does not include commercial leased line connectivity, which may affect the figures for very urban areas (e.g. city centres).
Comparability with WIMD 2014	New indicator

9. Housing domain

Conceptually, the purpose of the housing domain is to identify inadequate housing, in terms of physical and living conditions and availability. Here, living condition means the suitability of the housing for its inhabitant(s), for example in terms of health and safety, and necessary adaptations.

The domain has a relative weight of 7% in the overall index. This has increased from 5% in the 2014 index, due to the addition of a new modelled indicator on the likelihood of poor quality housing.

Domain construction

The income domain is made up of two equally weighted indicators:

- An indicator on overcrowding measures the percentage of people living in overcrowded households (2011 Census-based bedroom measure).
- A new modelled indicator on poor quality housing. It measures the likelihood of housing being in disrepair or containing serious hazards (for example, risk of falls or cold housing), and was calculated from a mixture of survey and administrative data sources by the Building Research Establishment (BRE).

For the two indicators, each LSOA was ranked in order, with the most deprived LSOA ranked 1 and the least deprived LSOA ranked 1,909. These ranks were assigned to a normal distribution (with low ranks receiving a low normalised value) before being combined with equal weighting. As with all domains, the final domain ranks were exponentially transformed, to form domain scores for use in the calculation of the overall WIMD 2019.

Methodological changes

The housing domain previously included two 2011 Census indicators:

- The percentage of the population living in overcrowded households
- The percentage of people living in a household with no central heating.

In WIMD 2019, the housing domain continues to include the indicator on overcrowding (to capture availability and to some extent living condition), and introduces a new modelled indicator on poor quality housing. This indicator helps to capture both physical and living conditions. For further information on why we dropped the indicator on lack of central heating, see our consultation report.

Indicators

Publication	WIMD 2019
Domain	Housing
Indicator Name	Percentage of people living in overcrowded households (bedrooms measure)
Source	2011 Census, Office for National Statistics (ONS)
Type of Indicator	Percentage
Denominator	Number of people living in households
Time period for WIMD 2019	2011
Additional Notes	This indicator provides a measure of whether a household's accommodation is overcrowded (based on the number of bedrooms). The ages of the household members and their relationships to each other are used to derive the number of bedrooms they require, based on a standard formula. The number of bedrooms required is subtracted from the number of bedrooms in the household's accommodation to obtain the occupancy rating. An occupancy rating of -1 implies that a household has one fewer bedroom than required. A bedroom is defined as any room that was intended to be used as a bedroom when the property was built, or any room that has been permanently converted for use as a bedroom. It also includes all rooms intended for use as a bedroom even if not being used as a bedroom at the time of the census. Bedsits and studio flats are counted as having one bedroom. For further information see the ONS quality notes on Census methodology.
Comparability with WIMD 2014	The same indicator is used for WIMD 2019 as for WIMD 2014

Publication	WIMD 2019
Domain	Housing
Indicator Name	Likelihood of poor quality housing (being in disrepair or containing serious hazards)
Source	Building Research Establishment, using various survey and administrative data sources (more detail in Annex H)
Type of Indicator	Percentage (likelihood)
Denominator	Numbers of residential dwellings
Time period for WIMD 2019	2017/18
Additional Notes	This is the first instance of using modelled data in any domain of WIMD, and has improved the range of information on housing available at the local area. However one drawback of using modelled data on an ongoing basis is the possible limitation in reflecting the impact of recent interventions or other changes. If informing decisions, modelled data should be used alongside robust, up-to-date local data or intelligence wherever possible. The new indicator is calculated using a model built from survey data, which makes probabilistic predictions about individual level dwellings in Wales, using a range of administrative datasets as inputs. This allows us to estimate the likelihood that dwellings in a given area: - contain a Category 1 hazard for excess cold, falls or other hazards under the Housing Health and Safety Rating System (HHSRS) - or are in a state of disrepair A dwelling is determined to have a Category 1 hazard as a result of excess cold if there is a severe threat from sub-optimal indoor temperatures. A dwelling is assessed as having a Category 1 hazard in terms of falls if there is determined to be a serious risk of falling on stairs, between levels, level surfaces or falling associated with a bath, shower or similar facility. A dwelling is said to be in disrepair if at least one of the key building components is old and needs replacing or major repair due to its condition; or more than one of the other building components are old and need replacing or major repair. Further details on the modelling process are provided in Annex H
Comparability with WIMD 2014	New indicator

10. Community Safety domain

The purpose of this domain is to measure deprivation relating to living in a safe community. It covers actual experience of crime and fire, as well as perceptions of safety whilst out and about in the local area.

The domain has a relative weight of 5% in the overall index.

Domain construction

There are six indicators in the community safety domain, weighted as follows. Factor analysis was used to calculate the indicator weights.

- 61% Police Recorded Criminal Damage
- 19% Police Recorded Violent Crime
- 8% Police Recorded Anti-Social Behaviour
- 5% Police Recorded Burglary
- 4% Police Recorded Theft
- 3% Fire Incidents

All of the indicators are expressed as rates, with the numerator consisting of the number of crimes/incidents/fires per LSOA (averaged over two years) and the denominator consisting of the number of people/properties in the LSOA.

For every indicator, each LSOA was ranked in order, with the most deprived LSOA ranked 1 and the least deprived LSOA ranked 1,909. These ranks were assigned to a normal distribution, with low ranks receiving a low normalised score. They were then combined using the weights above. As with all domains, the final domain ranks were exponentially transformed to form a domain score (see annex F).

Methodological changes

Minor changes have been made to the methodology for constructing the crime and anti-social behaviour incident numerators. These changes include the adoption of the same geographical aggregation technique and constraining technique as used in the Crime Domain of the English Indices of Deprivation 2019. The two-stage quality assurance process with the police forces also mirrors that adopted in the Crime Domain of the English Indices of Deprivation 2019. See following section for further details.

Small changes to the detailed categories composing each of the recorded crime indicators were also necessitated due to changes to the Home Office Counting Rules since the publication of the WIMD 2014. For example, burglaries that would previously have been coded as 28A/B/C/D (‘burglary in a dwelling’) are now coded as 28E/F/G/H (‘burglary residential’).

Prior to July 2013, crime codes 9A ('public fear, alarm or distress') and 9B ('racially or religiously aggravated public fear, alarm or distress') were classified as 'Violence against the person' but, following consultation, they have been reclassified as 'Public order offences'. These codes were included in the 'violent crime' indicator for WIMD 2014 but they have been excluded for WIMD 2019.

Methodology – Police Recorded Crime/Anti-Social Behaviour indicators

Data on recorded crimes and incidents of anti-social behaviour were made available at microdata level by the four police forces in Wales (Dyfed Powys, Gwent, North Wales and South Wales). The datasets contained information on the type of crime/incident, the date the crime/incident was recorded, and the geographical location of occurrence (i.e. postcode and/or grid reference). The data covered the two operational years of 2017-18 and 2018-19. Due to the sensitive nature of this information, all data processing was undertaken within a secure setting at a police force headquarters.

The burglary indicator consists of all categories of burglary recorded by the police (including domestic and non-domestic burglaries). Likewise, the criminal damage indicator contains all categories of criminal damage. In contrast, the violence and theft indicators are based on selected subsets of categories of these two crime types. For the anti-social behaviour indicator, all three categories of incident are included. Further details on the detailed crime and incident categories are provided for each of the indicators in the text below.

The methodology for producing the numerators for the indicators of crime and anti-social behaviour is broadly comparable to that used in the English Indices of Deprivation 2019:

- Firstly the police microdata was coded to ensure that every record was assigned a crime/incident type according to the Home Office Counting Rules. The crime types that were not needed for the community safety domain were removed from the microdata.
- The contents of the microdata were quality-assured by comparing the numbers of records (separately by crime/incident type and operational year) against the aggregate crime/incident statistics at police force level, as provided by the Home Office.
- The microdata was aggregated to LSOA level using the same approach as developed for and applied in the crime domain of the English Indices of Deprivation 2019. This aggregation technique introduces a degree of spatial smoothing to account for known issues with the police data geocoding. Specifically, a 100m buffer was drawn around a crime/incident location and that crime/incident was shared equally by all LSOAs intersecting with the 100m buffer. This process is undertaken separately for every crime and incident record in the base microdata. The output from the geographical aggregate stage is a set of five 'composite indicators', relating to 'violence', 'burglary', 'theft', 'criminal damage' and 'anti-social behaviour'.

- The outputs of the geographical aggregation stage were then quality-assured by comparing the mapped crime/incident counts against the aggregate crime/incident statistics at police force and community safety partnership level, as provided by the Home Office. Through continued close engagement with the four police forces, it was ensured that the LSOA level mapped crime/incident counts represented a comprehensive account of recorded crime and anti-social behaviour over the periods 2017-18 and 2018-19.
- Finally, the LSOA-level mapped crime/incident counts were constrained to the aggregate crime/incident statistics at police force and community safety partnership level, as provided by the Home Office. The purpose of this step was to deal with any residual differences between the four police forces in the proportion of crimes/incidents successfully mapped to LSOA level. The final LSOA level crime/incident counts therefore summed exactly to the aggregate crime/incident statistics at police force and community safety partnership level, as provided by the Home Office.

Indicators

Publication	WIMD 2019
Domain	Community Safety
Indicator Name	Police Recorded Burglary
Source	Welsh Police Forces
Type of Indicator	Rate (per 100)
Denominator	The average of 2018 and 2019: numbers of residential and business dwellings
Time period for WIMD 2019	2017-18 and 2018-19
Additional Notes	This indicator is the number of recorded burglary offences per 100 properties at LSOA level. The numerator is based on individual event-level recorded crime data provided by the four police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent). Counts of crime at LSOA level are constructed by aggregating the crimes using the postcodes and/or grid references contained within the data. The definition of burglary used here encompasses crimes against both residential and non-residential properties. The crime codes (as defined by the Home Office Counting Rules) included were: • 28E/F/G/H: Burglary residential • 29A: Aggravated Burglary residential • 30C/D: Burglary business and community • 31A: Aggravated burglary business and community More information on the Home Office Counting Rules can be found at: https://www.gov.uk/government/publications/counting-rules-for-recorded-crime
Comparability with WIMD 2014	Broadly comparable

Publication	WIMD 2019
Domain	Community Safety
Indicator Name	Police Recorded Theft
Source	Welsh Police Forces
Type of Indicator	Rate (per 100)
Denominator	Total resident population (2017 Small Area Population Estimates), excluding total 2017 prison population
Time period for WIMD 2019	2017-18 and 2018-19
Additional Notes	This indicator is the number of recorded theft offences per 100 people at LSOA level. The numerator is based on individual event-level recorded crime data provided by the four police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent). Counts of crime at LSOA level are constructed by aggregating the crimes using the postcodes and/or grid references contained within the data. The crime codes (as defined by the Home Office Counting Rules) included were: • 37/2 Aggravated Vehicle Taking • 39 Theft from the Person • 45 Theft from a Vehicle • 48 Theft or Unauthorised Taking of a Motor Vehicle and • 126 Vehicle Interference and Tampering. More information on the Home Office Counting Rules can be found at: https://www.gov.uk/government/publications/counting-rules-for-recorded-crime
Comparability with WIMD 2014	Broadly comparable

Publication	WIMD 2019
Domain	Community Safety
Indicator Name	Police Recorded Criminal Damage
Source	Welsh Police Forces
Type of Indicator	Rate (per 100)
Denominator	Total resident population (2017 Small Area Population Estimates), excluding total 2017 prison population
Time period for WIMD 2019	2017-18 and 2018-19
Additional Notes	This indicator is the number of recorded criminal damage offences per 100 people at LSOA level. The numerator is based on individual event-level recorded crime data provided by the four police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent). Counts of crime at LSOA level are constructed by aggregating the crimes using the postcodes and/or grid references contained within the data. The crime codes (as defined by the Home Office Counting Rules) included were: • 56A Arson endangering life • 56B Arson not endangering life • 58A Criminal Damage to a Dwelling • 58B Criminal Damage to a Building other than a Dwelling • 58C Criminal Damage to a Vehicle • 58D Other Criminal Damage • 58J Racially or Religiously Aggravated Criminal Damage • 59 Threat or possession with intent to commit Criminal Damage More information on the Home Office Counting Rules.
Comparability with WIMD 2014	Broadly comparable

Publication	WIMD 2019
Domain	Community Safety
Indicator Name	Police Recorded Violent Crime
Source	Welsh Police Forces
Type of Indicator	Rate (per 100)
Denominator	Total resident population (2017 Small Area Population Estimates), excluding total 2017 prison population
Time period for WIMD 2019	2017-18 and 2018-19

Additional Notes	This indicator is the number of recorded violence offences per 100 people at LSOA level. The numerator is based on individual event-level recorded crime data provided by the four police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent). Counts of crime at LSOA level are constructed by aggregating the crimes using the postcodes and/or grid references contained within the data. The crime codes (as defined by the Home Office Counting Rules) to be included were: • 1 Murder • 2 Attempted Murder • 3A Conspiracy to commit murder • 3B Threats to kill • 4/1 Manslaughter • 4/2 Infanticide • 5D Assault with Intent to Cause Serious Harm • 5E Endangering Life • 8L Harassment • 8M Racially or Religiously Aggravated Harassment • 8N Assault with Injury • 8P Racially or Religiously Aggravated Assault with injury • 8R Malicious Communications • 8S Assault with injury on a constable • 10A Possession of Firearms with intent • 10B Possession of Firearms Offences • 11A Cruelty to Children / Young Persons • 34A Robbery of Business Property • 34B Robbery of Personal Property • 37/1 Causing Death by Aggravated Vehicle Taking • 81 Firearms Act 1968 and other Firearms Act • 105A Assault without injury • 105B Racially or Religiously Aggravated Assault without injury More information on the Home Office Counting Rules.
Comparability with WIMD 2014	Broadly comparable

Publication	WIMD 2019
Domain	Community Safety
Indicator Name	Anti-Social Behaviour
Source	Welsh Police Forces
Type of Indicator	Rate (per 100)
Denominator	Total resident population (2017 Small Area Population Estimates), excluding total 2017 prison population
Time period for WIMD 2019	2017-18 and 2018-19
Additional Notes	This indicator is the number of reported anti-social behaviour incidents per 100 people at LSOA level. The numerator is based on individual event-level anti-social behaviour incident data provided by the four police forces in Wales (North Wales, Dyfed Powys, South Wales and Gwent). Counts of incidents at LSOA level are constructed by aggregating the incidents using the postcodes and/or grid references contained within the data.
Comparability with WIMD 2014	Broadly comparable

Publication	WIMD 2019
Domain	Community Safety
Indicator Name	Fire Incidences
Source	Incident Recording System (IRS), Welsh Government
Type of Indicator	Rate (per 100)
Denominator	Total resident population (2017 Small Area Population Estimates), excluding total 2017 prison population
Time period for WIMD 2019	2017-18 and 2018-19
Additional Notes	This indicator captures actual experiences of fire. Incidents of all primary, secondary and chimney fires were collected as counts by LSOA. Incidents requiring call out of fire and rescue services are related to deprivation and more likely within disadvantaged groups. Primary fires comprise 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. Chimney fires are reportable fires in occupied buildings where the fire was confined within the chimney structure and did not involve casualties or rescues or was attended by 5 or more appliances.
Comparability with WIMD 2014	Broadly comparable

11. Physical Environment domain

The purpose of this domain is to measure factors in the local area that may impact on the wellbeing or quality of life of those living in an area.

The domain has a relative weight of 5% in the overall index.

Domain construction

The physical environment domain is made up of three sub-domains weighted as follows:

- 40% Air Quality (three indicators)
- 40% Flood Risk (one indicator)
- 20% Green Space (two indicators)

Methodological changes

There have been several methodological changes to the physical environment domain between WIMD 2014 and WIMD 2019. A full list of the changes is outlined below.

In WIMD 2014, the Air Quality sub-domain comprised two indicators – Air Concentrations and Air Emissions – that were calculated using a combination of concentration data for a range of pollutants and Air Quality Management Areas. The sub-domain has been simplified for WIMD 2019 and now comprises three indicators based on the population weighted average concentration values of the following key pollutants:

- Nitrogen dioxide (NO_2)
- Particulates < 10 µm (PM_{10})
- Particulates < 2.5 µm ($PM_{2.5}$)

A Green Space sub-domain has been included in WIMD 2019 instead of the Proximity to Waste Disposal and Industrial Sites sub-domain. This sub-domain is comprised of two indicators:

- Proximity to accessible, natural green space – measuring the proportion of households within 300 metres of an accessible, natural green space, and
- Ambient green space score – measuring the mean household Normalised Difference Vegetation Index (NDVI).

The Flood Risk sub-domain for WIMD 2019 has been sourced from the Flood Risk Assessment Wales (FRAW) dataset developed and provided by Natural Resources Wales (NRW). This data includes information on flood risk from rivers, the sea and surface water flooding. This is a change from WIMD 2014 where the Flood Risk sub-domain calculation did not include surface water flooding.

Air Quality indicators

The Air Quality sub-domain comprises three separate indicators measuring concentrations of key pollutants. They are created using measurements of pollutants that could have negative effects on

human health and/or the environment, based on the best medical and scientific understanding, and are proposed as a 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.

Publication	WIMD 2019
Domain	Physical Environment
Indicator Names	Population Weighted Average Concentration Value – NO_2 Population Weighted Average Concentration Value – PM_{10} Population Weighted Average Concentration Value – $PM_{2.5}$
Source	Department for Environment, Food & Rural Affairs (DEFRA), Small Area Population Estimates (SAPE) – Office for National Statistics
Type of Indicator	Values
Denominator	n/a
Time period for WIMD 2019	2017
Additional Notes	Each year the UK Government's Pollution Climate Mapping (PCM) model calculates average pollutant concentrations for each square kilometre of the UK. The model is calibrated against measurements taken from the UK's national air quality monitoring network. To calculate indicators for WIMD 2019, this published data is used to assign a concentration of NO_2, PM_{10} and $PM_{2.5}$ to each residential dwelling in Wales based on which square kilometre of Wales it sits in. For each LSOA, the pollutant concentrations associated with each dwelling within it were averaged to give average NO_2, PM_{10} and $PM_{2.5}$ concentration values. These values were then population weighted.
Comparability with WIMD 2014	Not comparable

Air Quality sub-domain

To calculate the overall Air Quality sub-domain score, each indicator value was adjusted (via transformation) using a factor based on the objective, standard or risk factor for that specific pollutant and statistic. This method was developed to take into account air quality standards for each substance, which are based on the best medical and scientific understanding of their effects on health and/or the environment. The method also ensures that areas which have high prevalence of certain pollutants, but not others, are ranked as highly deprived; low levels of one pollutant will not cancel out the effect of a high level of another pollutant.

The transformed indicator values for each LSOA were averaged and ranked. These ranks were then exponentially transformed to produce a sub-domain score for each LSOA.

Flood Risk indicator and sub domain

Publication	WIMD 2019
Domain	Physical Environment
Indicator Name	Flood Risk
Source	Flood Risk Assessment Wales (FRAW) data - Natural Resources Wales (NRW)
Type of Indicator	Score (between 1 and 100)
Denominator	n/a
Time period for WIMD 2019	2017
Additional Notes	The Flood Risk indicator considers the proportion of households at risk of flooding from rivers, the sea and surface water flooding. The risk is based on predicted frequency, rather than the level of damage caused by flooding. The risk categories used are as follows: • Low risk - Less than 1 in 100 (1%) chance in any given year • Medium risk - Less than 1 in 30 (3.3%) but greater than or equal to 1 in 100 (1%) chance in any given year • High risk - greater than or equal to 1 in 30 (3.3%) chance of flooding in any given year To ensure the areas at risk of more severe flooding rank as more deprived than areas at risk of less severe flooding, the following weighting was given: • The number of households in an area at high risk was multiplied by 24; • The number of households in an area at medium risk was multiplied by 4; and • The number of households in an area at low risk was multiplied by 1. More information on the methodology for deriving the above weights is available at Annex I. Note that, in cases where household were at different levels of risk from different types of flooding, the highest risk level was given priority. Each of these numbers is calculated for each LSOA and then added together to give total normalised number of households at a risk of flooding per LSOA. This number is then divided by the total number of households in the LSOA to give the proportion of households at risk of flooding. These values are then ranked and exponentially transformed to produce an overall Flood Risk score.
Comparability with WIMD 2014	Not comparable

Green space indicators

The Green Space sub-domain comprises two indicators which measure the proximity to accessible, natural green space and the amount of ambient green space respectively.

Publication	WIMD 2019
Domain	Physical Environment
Indicator Name	Proximity to Accessible, Natural Green Space
Source	OS MasterMap Topography Layer®, AddressBase® Plus
Type of Indicator	Percentage
Denominator	Number of households in LSOA
Time period for WIMD 2019	2019
Additional Notes	This indicator measures the proportion of households in each LSOA that are within a 5 minute walk / 300 metres of an accessible natural green space. To calculate this indicator, Greenspace footprints were derived from OS MasterMap Topography Layer®, scope defined by OS Open Greenspace combined with Natural Resources Wales' recognised natural greenspace typologies. The output highlights sites that could confidently be described as natural feeling places to which the public have right of access. Sites such as golf courses, allotments and cemeteries were excluded from the list. To approximate a 5 minute walk, time-travel polygons with a radius of 300 metres around the green space sites were calculated. All residential dwellings (sourced from AddressBase® Plus) were then intersected with these polygons and flagged if they were within 300 metres of an accessible, natural green space. Counts of dwellings were then aggregated to an LSOA level to calculate the proportion of dwellings within a 5 minute walk / 300 metres of an accessible, natural green space. Further detail on the derivation of this indicator can be found in Annex J.
Comparability with WIMD 2014	New indicator

Publication	WIMD 2019
Domain	Physical Environment
Indicator Name	Ambient Green Space Score
Source	ADR-Wales
Type of Indicator	Score
Denominator	n/a
Time period for WIMD 2019	2019 - Aerial imagery captured over revisional time series
Additional Notes	This indicator measures the ambient greenness within each LSOA. It is calculated as the Mean Normalised Difference Vegetation Index (NDVI) within a 300 metre Euclidean buffer around each residential dwelling. Euclidean buffers are not bound to the boundary of an LSOA as it is recognised that an LSOA's geography does not necessarily represent human behaviour. This removes any edging effect whereby a household may be located within a particular LSOA but their Euclidean buffer overlaps another. The NDVI calculates the normalised difference between the red and infrared bands for quantitative and standardised measurement of vegetation presence and health. Healthy vegetation has a high reflectance of Near-Infrared wavelengths and greater absorption of red wavelengths due to a greater chlorophyll composition. The Near-Infrared (NIR) and red (R) spectral channels of an image are used to calculate an index value, using the following equation: $$NDVI = \frac{(NIR - R)}{(NIR + R)}$$ To calculate NDVI values for this indicator, WISERD, on behalf of Welsh Government sourced 50cm Colour Infrared imagery from Aerial Photography Great Britain Consortium (Bluesky International Limited and Getmapping Plc) via the Public-Sector Mapping Agreement (PSMA). Residential dwellings were sourced from AddressBase® Plus.
Comparability with WIMD 2014	New indicator

Green Space sub-domain

To calculate the overall Green Space sub-domain score, each set of indicator values was ranked and normalised. The normalised values were then combined using the following weighting:

- 60% Proximity to Accessible, Natural Green Space
- 40% Ambient Green Space Score

The combined values were then re-ranked and exponentially transformed to produce a sub-domain score for each LSOA.

Annex A: Exponential Transformation of the Domain Indexes

The exponential transformation of ranks reduces the extent to which deprivation in some domains can be cancelled by lack of deprivation in others. The transformation 'draws out' the ranks of the most deprived areas so that spaces are introduced between areas that reflect the actual distributions, and emphasise the most deprived 'tail' of the distribution.

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=1/1909 for the least deprived, R=1909/1909=1 for the most deprived). The transformed domain, (X) equals:

$-23 \times \log\{1 - R \times [1 - \exp(-100/23)]\}$,

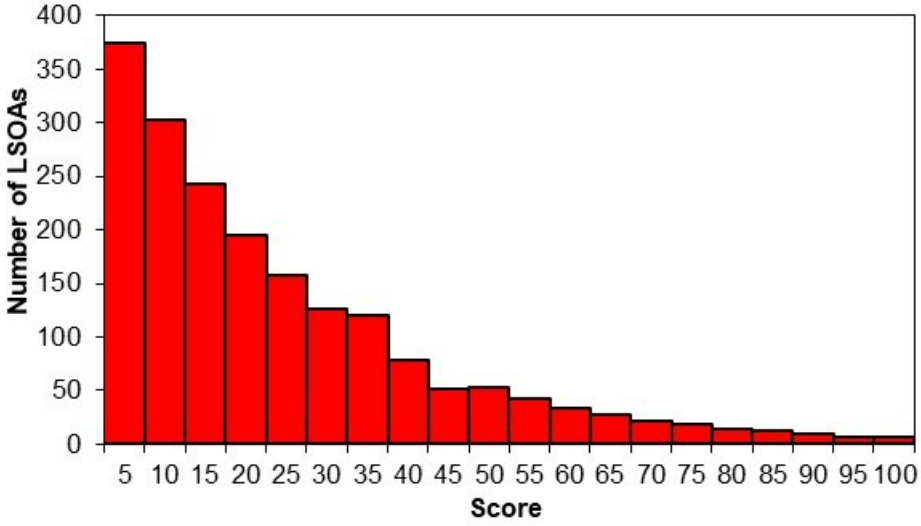
where log denotes natural logarithm and exp the exponential or antilog transformation. This formula is straightforward to calculate and is in fact simpler than the commonly-used transformation to a normal curve, which requires the use of a look-up table. The resulting distribution is illustrated below.

Each transformed domain has a range of 0 to 100, with a score of 100 for the most deprived LSOA. The constant -23 gives a 10% cancellation property. This means that ten% of LSOAs have a score higher than 50 (the most deprived), and the remaining 90% of LSOAs have scores between 0 and 50.

When transformed scores from different domains are combined by averaging them, the skewness of the distribution reduces the extent to which deprivation in one domain can be cancelled by lack of deprivation in 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 at the 50th percentile; the high deprivation in one domain would have been fully cancelled by the low deprivation in the other.

This means the Index methodology is designed to reliably distinguish between areas at the most deprived end of the distribution, but differences between the least deprived areas in Wales are less well defined.

Histogram of a transformed domain



Annex B: 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, Access to Services and Community Safety. Factor analysis is used in these domains because they contain indicators that measure a number of forms of that deprivation, on potentially different metrics and with different levels of accuracy, 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; and
- 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, where forms of deprivation tend to present themselves in different ways at different times. 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. For example, in the Education domain, lack of qualifications in the adult population as well as poor results at school level were assessed. 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 illness, 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 2019 follows the methodology of recent iterations and that applied by Oxford University for WIMD 2000, as well as the Indices 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. 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 among which the choice 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 communality 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 2019 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 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; and
- 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.

Annex C: Allocation of Data to Lower layer Super Output Areas

For most domains, indicator data are allocated to LSOAs by the data suppliers as part of the collection process. However, data is provided at a lower geographical level for the travel time indicators in the Access to Services domain and for all indicators in the Education and Community Safety domains. An explanation of how indicator data in these domains were allocated to LSOAs is provided below.

Education domain

All new and updated indicator data in the Education domain was provided at the postcode level. As many postcodes do not sit wholly within one LSOA, a method of apportioning postcode-level data to multiple LSOAs was used.

In this apportionment method, data are weighted by the proportion of the dwellings in a postcode that sit within each intersecting LSOA. It is assumed that the proportions of dwellings in each intersecting LSOA will be broadly equivalent to the proportions of the postcode population living in each intersecting LSOA. Finally, where postcodes do sit wholly within one LSOA, data are given a weighting of 1.

To apply this method, a postcode-to-LSOA lookup (with calculated dwelling weights) was developed by the Welsh Government Geography and Technology Unit and initially used to match as many postcodes to LSOAs as possible.

Any non-matched postcodes were then subsequently matched using an ONS lookup that assigned postcodes to LSOAs based on the nearest geographical centre of an LSOA. This lookup was not weighted so a default weight of 1 was assigned to all matched data.

Any data still unmatched at this point was excluded from indicator calculations.

Access to Services domain

Service point location information (e.g. for schools, post offices etc.) used in the travel time indicators were geocoded and allocated to LSOAs using a Graphic Information System (GIS).

Community Safety domain

Data on recorded crimes and incidents of anti-social behaviour were made available at microdata level by the four police forces in Wales. These datasets included information on the geographical location of occurrence (i.e. postcode and/or grid reference) which were assigned to LSOAs using a spatial smoothing method. Specifically, a 100m buffer was drawn around a crime/incident location and that crime/incident was shared equally by all LSOAs intersecting with the 100m buffer. This process was undertaken separately for every crime and incident record in the base microdata to account for known issues with the police data geocoding

Data on the grid references of fire incidents were sourced from the Incident Recording System (IRS) and mapped to LSOAs by the Welsh Government Geography and Technology Unit.

Annex D: Health domain – Indirect Age-Sex Standardisation

Indirect standardisation involves applying age-sex specific rates observed at national level to the population structure of each LSOA. The reason for using age-sex standardisation for the WIMD health indicators is to adjust the indicators to allow for different age and sex distributions amongst LSOA populations. For example, one might expect to observe a higher rate of deaths in an aging population than in one consisting predominantly of young families. Standardisation attempts to adjust for these differences in population.

The number of expected incidences (for WIMD these are limiting long-term illness, cancer, chronic and mental health conditions, and premature death) for each age-sex group in an LSOA is estimated by multiplying the number of people in the given age-sex group in the LSOA by the age-sex specific rate observed for Wales as a whole for that age-sex group. The total number of expected incidences for the LSOA is calculated by totalling the number of expected incidences for each age-sex group. The standardised ratio (e.g. of cancer incidence) for each LSOA is the number of observed incidences in the LSOA divided by the number of expected incidences.

$$\text{standardised ratio} = \frac{\text{observed incidences}}{\text{expected incidences}}$$

An indirectly age-sex standardised rate can be obtained by multiplying the standardised ratio for the LSOA by the crude rate for all of Wales. The Welsh crude rate is the number of incidences observed in Wales divided by the total Welsh population. The result is expressed as a rate per 100,000 people.

Indirectly standardised rate = standardised ratio x Welsh crude rate x 100,000

Annex E: Further detail on the new indicators on GP-diagnosed health conditions

Background

The national Quality and Outcomes Framework (QOF) was introduced as part of the new General Medical Services (GMS) contract on 1 April 2004. As part of this framework, GP practices are required to maintain 'disease registers'; these are: lists of patients registered with the GP who have been diagnosed with the disease or risk factor described in the register indicator. While it is recognised that these may not be completely accurate, it is the responsibility of the contractor to demonstrate that it has systems in place to maintain a high quality register.

The exact terms of QOF are reviewed every year and this has resulted in new diseases registers being added to the QOF requirements, and some disease registers being removed from the QOF requirements. When a disease register is removed from the QOF requirements, GPs will still maintain these registers, but they are not incentivised to do so. As well as overarching disease registers, GPs can use sub-indicators called 'read codes' to record patients' specific conditions.

As a result of contract reform in 2019, QOF will be replaced by a new Quality Assurance and Improvement Framework (QAIF). This new framework will still require GP practices to maintain disease registers. Full guidance on QOF in Wales can be found on the NHS Wales website.

Part of the function of NHS Wales Informatics Service (NWIS) is to request and extract disease register data from GP practices. The Welsh Government was granted access to this data at small area level, based on patient residence, for the purpose of informing WIMD.

Data specification and collection

NWIS provided us with counts of patients with current diagnoses (as at 31 March 2019) for one or more selected conditions, separately for mental health and chronic health conditions. This included patients who had a diagnosis at any time period prior to 31 March 2019, as long as they were still on the register.

The counts were based on a defined list of disease registers and sub-indicators, see below. Data was received from all GP practices active in Wales, then aggregated to the Welsh Lower Layer Super Output Areas (LSOAs), according to patient's home address.

To compile a specification for the conditions to include, we began with the list of registers included on the QOF framework which we have published data on in recent years. This provides a timely, well-documented and established framework for the data. When deciding what to include in the WIMD indicators, we considered whether conditions were:

- less able to be managed by controls or treatment which allow the individual to lead a normal life, and
- more likely to cause substantial pain and severe disability, and are associated with decreased life expectancy.

The final lists of selected conditions follow below.

Mental health conditions

- Depression
- Low mood (patients with record of low mood and an active repeat prescription for an anti-depressant)
- Anxiety disorder (including panic disorders)
- Dementia
- Severe mental illnesses (schizophrenia, bipolar affective disorder and other psychoses)

Chronic health conditions

- Coronary Heart Disease
- Chronic Obstructive Pulmonary Disease
- Stroke and Transient Ischaemic Attack
- Peripheral Arterial Disease
- Chronic Kidney Disease
- Diabetes Mellitus (for 0-16 year olds this only included diagnosis for Type 1 diabetes, due to suspected under-reporting of other types of diabetes for children; for those aged 17+ this includes any diagnosis for diabetes)
- Epilepsy

Other conditions previously included on the QOF framework and sub-indicators (not previously including in QOF) were considered and discounted. This involved consideration of our criteria, of data quality and validation, and any suspected variation in diagnosis-rates by deprivation, in consultation with an actively practicing GP and the WIMD health domain group of experts. Some of the conditions considered but not counted included chronic liver disease and osteoarthritis, since we were not aware of any pre-existing agreed lists of codes representing these conditions, to use for data extraction. The list of disease registers and sub-indicators will be comprehensively reviewed and may result in additional conditions being included in future Indexes.

Data quality

The way in which certain conditions are recorded across GP surgeries may vary. How promptly patients are removed from registers when conditions are resolved may also vary between practices, and impact on our data. Broadly, a patient will be removed from the disease register by a GP when they are determined not to have the conditions anymore. However, many of the conditions in both the chronic and mental health conditions indicators are long-term and are unlikely to fully resolve.

Given the possible variation in recording practices, we analysed the data carefully to identify any possible systematic bias. Our quality assurance of the data involved:

- Data analysis, comparing against the WIMD 2014 health domain and overall deprivation ranks, analysis by local authority and local health board. This revealed unsurprising patterns, with high correlation between the rates of diagnosis, health deprivation and overall deprivation.
- Comparison with related data sources: the National Survey for Wales and prescriptions data. No obvious issues were identified, however differences in definitions and collection methods meant this exercise was limited in its usefulness.
- Comparisons with GP level QOF disease register data which did not highlight any concerns of data quality when using disease registers at low-level geographies.
- Discussion of results with health analysts, government officials, medical advisers, and local authority analysts.

To summarise, GP-recorded rates of diagnoses have not been analysed on a small area resident basis before, and there is no independent source to assure the quality of the information against. There is likely to be variation in GP recording practices. Some areas have high absolute rates (over 1 in 3) of diagnosed chronic or mental health conditions, and care should be taken in interpreting this data. However for the WIMD domain and overall Index ranks, it is only the relative rank of areas on the indicator that matters. Since comparison of the ranks with previous WIMD ranks showed high correlation, as expected, no immediate areas of concern were identified. Working towards the next WIMD, we will seek further sources of data to quality assure these indicators, and are open to feedback on their robustness and usefulness.

Denominator (population) data

The indicators within the health domain of WIMD are indirectly age-sex standardised to adjust for the expected prevalence of disease within the underlying population. This allows the Index to identify areas where health deprivations exists beyond the effect of age and sex.

For the denominator, we have used the latest available Small Area Population Estimates (mid-2018), minus the prison population (2018). This allows for the breakdowns required to standardise rates for the effect of different age and sex profiles in different areas. It is also consistent with denominators used in the income, employment and community safety deprivation domains of WIMD.

Another option might have been to use the patient register data, however the data available to use would not allow us to undertake age-sex standardisation using this source. Also, the ONS have published an assessment of the quality of the NHS patient register data. This explains that the patient register has a number of issues when used for statistical purposes. The source has a

number of both under- and over-coverage issues, limited audit and potential for distortive effects because of its role in GP finance. The effect of these issues will vary by geography, age and sex.

Data adjustments

There are some Welsh residents with diagnosed health conditions who would not be captured in the data recorded by Welsh GPs, and we have adjusted for these in two ways:

- There are nearly 14,000 Welsh residents registered with primary care providers in England. We have adjusted the rates for the 25 small areas with over a 100 residents registered in England, which account for nearly 13,000 of the 14,000 people affected. The standardised rates were adjusted by scaling them up to reflect the volume of residents for whom we are missing data, since we do not have information on any diagnoses made across the border.
- Prisoners are likely to remain registered with a GP at their home address rather than a GP local to their prison for the duration of their sentence. Therefore we have subtracted prisoner numbers from the population estimates (denominator) before our calculation of standardised rates. This affects five small areas.

Annex F: Parameters for Calculation of Physical Access to Services Indicators

A public consultation was held to agree proposals for WIMD 2019; this covered which services to include and the inclusion of indicators based on public and private travel times. The indicators were agreed with the Access to Services domain group.

Public transport

For routing along a Public Transport network Welsh Government have worked with the Data Science Campus, Office for National Statistics to produce propeR. propeR is an open-source, multi-modal trip planner that provides an R interface to the Open Trip Planner engine. You can find more information in the GitHub repository.

propeR will be used to generate time-travel isochrones, each representing 5 minutes of multi-modal travel (walk, bus, coach or train) for a given indicator.

propeR calculates travel times using Dikstra's algorithm for transit and Contraction Hierarchies for non-transit journeys against OpenTripPlanner. The graph consumes a General Transit Feed Specification (GTFS). A General Transit Feed Specification, or a GTFS feed, is a file that allows public transit agencies to publish their transit data in a format that can be consumed by a wide variety of software applications. In it includes information surrounding the geometry of the stops, as well as the route information and times of each service.

Welsh Government would like to thank Traveline Cymru for providing transit feeds for bus and coach services across Wales. This data was received in a TransXChange format (similar to GTFS) so had to be translated. The Data Science Campus provided the WIMD team with translated TXC files in the format of GTFS, as explained in the propeR manual.

Data for rail journeys across Wales was sourced from the Rail Delivery Group and was downloaded in CIF format. Registration is required for access. The propeR manual also shows how to convert CIF to GTFS. The Data Science Campus provided the WIMD team with the converted rail data.

propeR uses the following assumptions:

- The default walking speed is 1.4 m/s. Walking speeds are irrespective of terrain and obstacles, but walking is not permitted on roads deemed non-pedestrian in OpenStreetMap (e.g. motorways).
- No maximum walking time is assigned to traversals. Instead a maximum walking distance is applied for the entire trip and is set to 1 kilometre. This equates to roughly 20 minutes of walking. This can be done in isolation to any public transport, on-route to a stop, between stops, and for the onwards journey to the service.

- Walking is more favourable for multi-transfer journeys of the same total duration (but walking must be less or equal to the maximum walking distance).
- Public transport can be provided in England or Wales, where cross border travel is required to a citizens' nearest service.
- Isochrones are produced in 5-minute intervals to and from service locations and are cut off at 90 minutes. Any residential addresses exceeding a 90-minute travel time to a service is automatically set to 90 minutes.

Private transport

Private transport travel times to the nearest access point for a given service were calculated using the pgRouting library within PostGIS.

Journey times were calculated from every service origin to every residential address within Wales. The shortest of all travel times per indicator (in decimal minutes) equals the shortest journey time to the nearest service.

The vehicular network was captured in the form of Ordnance Survey (OS) MasterMap Highways Network which Welsh Government access from OS under the Public Sector Mapping Agreement. Highways Network is the most complete, detailed and accurate navigable road network dataset for Great Britain.

For the purposes of Access to Services private transport calculations, OSMM Highways link geometry was broken up into smaller links to create a new node in front of every residential address at their closest point to the road network. This ensured time travel calculations per dwelling were not falsified with an additional distance to the nearest node in the standard link-node data structure.

Average vehicle speeds were applied to every road link within the network. The speed is based on Trafficmaster data sourced from Basemap. This data identifies the average speed travelled across all roads in Great Britain at different times of day. The average speed is calculated based on detailed historical speed information, which is collected annually by in vehicle telematics devices and mapped to each unique OS MasterMap RoadLink TOID. The average speed value is provided in both directions.

The speed database contains detailed average speed information which is calculated using over 135,000 vehicles with inbuilt trackers and GPS loggers and mixed-use vehicle fleet telematic data. The main type of vehicle is cars, but LGV's and HGVS are also used to calculate the average speed. The data is polled every 1 to 10 seconds and then attached to the Ordnance Survey road network.

Within any given month of data collection, there will be 99% coverage of average speed information for Motorways, A Roads and B roads and over 50% coverage of c class roads. For road links or time periods where there is no speed data available over the 12 months, these road links are infilled using average speed data from road links of similar characteristics to ensure 100% coverage of data.

Residential addresses

The location of residential addresses for Wales were captured using an agreed definition from the AddressBase® Plus dataset (Epoch 66), obtained under the Public Sector Mapping Agreement.

For the purposes of the Access to Services domain, a residential address is included in the analyses if:

- The address is an active record
- The address has a valid domestic council tax record or adheres to one of the following classifications:
    - Residential
    - Residential Dwelling
    - Caravan (and has a valid domestic council tax record)
    - Detached
    - Semi-Detached
    - Terraced
    - Self-Contained Flat (Includes Maisonette / Apartment)
    - House Boat (and has a valid domestic council tax record)
    - Sheltered Accommodation
    - House in Multiple Occupation
    - HMO Parent
    - HMO Bedsit / Other Non Self-Contained Accommodation

There are 1,452,370 residential dwellings in Wales in the origins dataset.

Physical Services

The services included are:

- Pharmacy
- Food shop
- General Practitioner (GP)
- Public library
- Post office
- Primary school
- Sports Facility (replacing Leisure Centre in WIMD 2014)
- Petrol station (private transport time only)
- Secondary school

Travel times

Travel times are used for both public and private (rather than distance).

The journey to a service from a residential dwelling is defined as the *journey to the service*.

The journey to a residential dwelling from a service is defined as *the journey to the residential dwelling.*

The journey to the service is a defined as *single journey*, as is the journey to residential dwelling. The combination of these two elements is defined as the *return journey*.

Travel times will be calculated for a one way single journey at a specific time. The maximum specified travel time for a single journey will be 90 minutes (1.5 hours).

Any travel times over 90 minutes will be given the value of 90 minutes.

Travel times are defined as the time taken to travel from a residential dwelling to the services; the quickest travel times to the individual services will be used.

The private travel time and public travel times may refer to different services of the same type (i.e. an individual using public transport may go to a different GP than an individual using private transport).

Travel times to services located 10-kilometres into England will be included.

Time periods

Public transport: three travel times were calculated for both the journey to the service and the journey to the residential dwelling. The average (mean), of the three journeys to the service times were added to the average of the three journeys to the residential dwelling times. Calculating the average of the three journeys in both directions reflects the service frequency for a particular journey.

Private transport: a singular journey from every service to all residential dwellings was calculated. For every residential address, the smallest travel time was therefore taken as the quickest travel time. This journey was doubled to represent a return journey to/from the service.

Note: This has the same effect as averaging the two values and then doubling to take into account the return journey. Doubling the single journey time as the journey to and from the dwelling will show no/very little difference by private transport. This wouldn’t be the case by public transport.

The times of travel by public transport journeys were:

Service	Modelling date	Arrive/ Depart	Time		
Primary School	12/03/2019	Arrive	8:30	8:45	9:00
Primary School	12/03/2019	Depart	15:30	16:00	16:30
Secondary School	12/03/2019	Arrive	8:30	8:45	9:00
Secondary School	12/03/2019	Depart	15:30	16:00	16:30
Food shop	12/03/2019	Arrive	9:00	10:00	11:00
Food shop	12/03/2019	Depart	11:00	12:00	13:00
GP	12/03/2019	Arrive	9:00	10:00	11:00
GP	12/03/2019	Depart	11:00	12:00	13:00
Pharmacy	12/03/2019	Arrive	9:00	10:00	11:00
Pharmacy	12/03/2019	Depart	11:00	12:00	13:00
Public Library	12/03/2019	Arrive	9:00	10:00	11:00
Public Library	12/03/2019	Depart	11:00	12:00	13:00
Post Office	12/03/2019	Arrive	9:00	10:00	11:00
Post Office	12/03/2019	Depart	11:00	12:00	13:00
Sports Facility	12/03/2019	Arrive	16:00	17:00	18:00
Sports Facility	12/03/2019	Depart	18:00	19:00	20:00

The times of travel by private transport journeys were:

Service	Weekday travel time period
Primary School	07:00 – 09:00 Monday – Friday (Peak AM)
Secondary School	07:00 – 09:00 Monday – Friday (Peak AM)
Food Shop	10:00 – 16:00 Monday – Friday (Off Peak)
GP	10:00 – 16:00 Monday – Friday (Off Peak)
Pharmacy	10:00 – 16:00 Monday – Friday (Off Peak)
Public Library	10:00 – 16:00 Monday – Friday (Off Peak)
Post Office	10:00 – 16:00 Monday – Friday (Off Peak)
Petrol Station	10:00 – 16:00 Monday – Friday (Off Peak)
Sports Facility	16:00 – 19:00 Monday – Friday (Peak PM)

Note: Petrol stations are only included in the private transport journey travel times.

Data sources

Where possible, the most authoritative data set has been sought for Wales. A citizen's ability to use any of the services included in this domain is independent of their geography (i.e. being a Welsh resident). This exercise, therefore, acknowledges that a citizens' nearest service may fall beyond the border of Wales. To accommodate this, service locations falling within a 10-kilometre buffer of the Welsh border were included. Consequently, some data sets have a substitute source to cover England in cases where the primary and preferred Welsh source is unavailable for that geography.

The following data sources were used for the locations of services.

Service	(Num. of) Locations	Data Source	Complied by	Currency
GPs	625	Medical Performers List (Wales)	NHS Wales Shared Services Partnership	March 2019
Primary Schools	1,266 (inc. 19 Middle Schools)	Welsh Government (Wal); AddressBase® Plus (Eng)	Welsh Government Education Team	March 2019
Secondary Schools	283 (inc. 19 Middle Schools)	Welsh Government (Wal); AddressBase® Plus (Eng)	Welsh Government Education Team	March 2019
Sports Facilities	353	Sport Wales (Wal); Sport England (Eng)	Welsh Government	March 2019
Food Shops	3,129	Points of Interest® (Wal + Eng)	Welsh Government	March 2019
Post offices	910	Points of Interest® (Wal + Eng)	Welsh Government	March 2019
Public Library	276	Local Authorities (Wales) Points of Interest® (Eng)	Welsh Government	March 2019
Pharmacies	817	All Wales Pharmacy Database (Wal); Points of Interest® (Eng)	NHS Wales Shared Services Partnership	March 2019
Petrol stations	695	Points of Interest® (Wales + Eng)	Welsh Government	March 2019

Data quality assurarce and validation

Data were quality assured by the Geography & Technology department, as well as the WIMD team.

At a raw data level – the following steps were taken by the Geography & Technology department to quality assure the data:

Public Transport

- Quality assurance of all UPRNs whose travel time exceeded 90 minutes.
- For each of the indicators, the entirety of Wales was qualitatively sense checked by visualising the travel times as a heat map.
- Randomised checks of isochrone geometries against bus and train timetable information to ensure services are accurately reflected and being modelled.

Private Transport

- Network level checks – sampling of the OS MasterMap Highways Network geometry and average speed data to ensure no anomalous links or obstructions with the routing graph.
- Quality assurance of all UPRNs whose travel time exceeded 90 minutes.
- For each of the indicators, the entirety of Wales was qualitatively sense checked by visualising the travel times as a heat map.
- A comparison of the results to Google Maps routing platform for a sub-set of services (against all indicators).

The WIMD team also undertook statistical quality assurance checks for both public and private travel times for each of the indicators, by checking LSOA travel times for WIMD 2019 relative to WIMD 2014. This involved:

- Sense-checking data by observing maximum, mean and median absolute changes in indicator times for LSOAs between 2014 and 2019.
- For each LSOA with a maximum absolute change in travel time, a comparison search was made using the Google Maps routing platform to verify the 2019 travel time.
- Sense-checking travel times in outlying LSOAs, and in those in which a rank had shifted by 7 or more deciles in either direction.
- Sense-checking the distribution of travel times across all LSOAs, local authorities, and settlement types.

Annex G: Methodology for the Digital Access to Services indicator

Ofcom collects and analyses data from the main fixed broadband internet service providers on addresses covered by their service. The Ofcom Connected Nations Update: Spring 2019 collected coverage data as a snapshot in January 2019. The Connected Nations 2018 Methodology Report (on which this 2019 spring update is based) explains how Operators were asked to provide data for each address where a service was provided. Due to variations in broadband performance over time, the file should not be regarded as a definitive and fixed view of fixed broadband infrastructure. However, the information provided is useful in identifying variations in broadband availability.

Coverage

For the overall coverage of fixed broadband Ofcom identified the number of residential and small business premises, through premises data from the Ordnance Survey (OS) Addressbase® Premium dataset (August 2018, Epoch 60). This is combined with additional geographic classifications from the ONS National Statistics Postcode Lookup (NSPL) (August 2018). The identification of premises is based on delivery point locations, excluding PO boxes and large organisations. Additionally, unless otherwise specified, the analysis only includes premises that are recorded as approved and constructed.

This approach is the same applied throughout the Connected Nations report series. In addition to the delivery point base, which is derived from the Royal Mail Postcode Address File® data set, premises may be identified by Local Authorities as an addressable location. These records are not counted in the premise base.

The availability of address-level data allows Ofcom to create a comprehensive data set describing the characteristics of all available services and operators present at premises across the UK. Many operators provided a unique property reference number (UPRN), a common identifier available for use in the UK. Other operators provided address information that is processed and linked to the Ofcom premise base. For premises not matched by UPRN or building address information, a postcode level estimate was applied where, providing that a postcode unit contained at least one matched premises, any non-matched premises were assigned the best available coverage.

Calculating availability

Each operator provides information on the technology available together with predictions of download and upload speeds. After the address matching process these characteristics are assigned to each premise to enable further detailed analysis to be undertaken. Ofcom apply thresholds in their analysis to investigate different patterns of provision. For coverage they use the maximum predicted download speed available at a premise to determine in which broadband category a premise is represented. Ofcom use 30 Mbit/s download speeds as the threshold for

defining superfast services. Additionally, Ofcom include all unmatched and unclassified premises as not having access to a decent broadband service.

Performance metrics, speeds and data use

Ofcom gather data from many of the fixed broadband internet service providers (see below) on both their retail services and the services they provide to other ISPs as a wholesale service. Ofcom analysis of broadband speeds is based on the information provided by these ISPs regarding the sync speed of each active line. This gives a measure of the maximum possible connection speed achievable between the ISP's access network and the consumer premises. Line speed measurements are typically a few Mbit/s lower than sync speed measurement, and they typically vary throughout the day depending on the level of congestion in the ISP's network.

Obtaining information from providers

Ofcom requested data from communication providers using powers under section 135 of the Communications Act 2003 and Regulation 17 of the Statutory Instrument 2016/607. Under section 134A and 134B of the Act Ofcom is required to prepare a report for "each relevant period", as defined in section 134A(4) of the Act, that deals with the electronic communications networks matters listed in section 134B(1), and the electronic communications services matters listed in section 134B(2), of the Act.

The following fixed network providers supplied data for use in the Ofcom Connected Nations report:

- B4RN
- Bridge Fibre
- BT Group
- Cablecom Glide
- Call Flow
- CityFibre
- Community Fibre
- Gigaclear
- Hutchinson 3G UK Limited ("Three")
- Hyperoptic
- IFNL
- ITS
- KCOM
- Openreach
- Sky
- Spectrum Internet
- TalkTalk
- Truespeed
- Virgin Media
- Vodafone
- VX Fibre
- Wessex Internet
- Wight Fibre

Annex H: Modelled poor quality housing indicator - Methodology

In 2019, Welsh Government commissioned the Building Research Establishment (BRE) to develop a poor quality housing indicator for Wales, at small area (LSOA) level. BRE used a similar methodology to that they employed for the English Index of Multiple Deprivation 2019. The new indicator is calculated using a model built from survey data, which makes probabilistic predictions about individual level dwellings in Wales, using a range of administrative datasets as inputs.

Developing the poor quality housing indicator

Conceptually, the purpose of the housing domain is to identify inadequate housing, in terms of physical and living conditions and availability. The poor quality housing indicator estimates the likelihood that dwellings in a given area:

- contain a Category 1 hazard for excess cold, falls or other hazards under the Housing Health and Safety Rating System (HHSRS)
- or are in a state of disrepair

These two measures – disrepair and hazards (HHSRS) – are useful and well-established indicators of housing deprivation, also used for the 2019 update of the English Index of Multiple Deprivation.

Summary of method

The approach that was used to create the modelled data involved:

- Using data from housing condition surveys where experienced surveyors carried out physical inspections of a sample of properties from all tenures (including measurement of disrepair and risks of hazards across all types of dwellings)
- Building a model from this data to predict the likelihood of poor quality housing based on a range of predictors (such as the age, type, size, tenure, construction and energy variables such as heating and fuel type)
- Applying this model to all dwellings in Wales, using data from a range of sources (including Ordnance Survey, Land Registry, EPC data) to provide the required predictors
- Benchmarking the results to national estimates of poor housing quality from the Welsh Housing Conditions Survey 2017-18

The modelling process above will be carried out separately for the aspects of poor quality housing listed below (and defined further in following sub-sections), at a dwelling level:

- The presence of a Category 1 hazard for Excess cold (using SAP ratings as a proxy measure[1])
- The presence of a Category 1 hazard for Falls
- The presence of a Category 1 hazard for a Hazard other than Excess cold or Falls
- Being in disrepair

A dwelling is classed as being of poor quality housing if it is predicted to have any of the above features. Estimated data for individual dwellings will then be aggregated to LSOA level to provide a rate of dwellings which are classed as poor quality housing for each LSOA.

[1] The BRE Model is based on SAP 2012 (version 9.92), and in this methodology a dwelling is considered to have a Category 1 Excess Cold hazard if it has a SAP score of less than 33.52.

Hazards

The Housing Health and Safety Rating System (HHSRS) is used by Welsh Government as an evidence based risk assessment procedure for residential properties. The HHSRS is a means of identifying defects in dwellings and of evaluating the potential effect of any defects on the health and safety of occupants, visitors, neighbours and passers-by. The system provides a means of rating the seriousness of any hazard so it is possible to differentiate between minor hazards and those where there is an imminent threat of major harm or even death. The emphasis is placed on the potential effect of any defects on the health and safety of occupants, visitors, and particularly vulnerable people. 29 hazards make up the HHSRS. For the purposes of this modelling, hazards were grouped into the following three categories:

1. Excess cold
2. Falls (comprising of falls on stairs, falls on the level, and falls associated with baths)
3. Other (all other hazards)

Dwellings where at least one Category 1 hazard (hazards with potential extreme harm outcome) was present were reported as failing the assessment. A dwelling is determined to have a Category 1 hazard as a result of excess cold if there is a severe threat from sub-optimal indoor temperatures. A dwelling is assessed as having a Category 1 hazard in terms of falls if there is determined to be a serious risk of falling on stairs, between levels, level surfaces or falling associated with a bath, shower or similar facility.

Disrepair

The same Disrepair criterion as used in the Decent Homes Standard (England) was used, (the same as that used for the 2019 update of the English Index of Multiple Deprivation). A dwelling failing the Disrepair criterion means that certain building components are in poor condition, defined as:

1. One or more key building components are old and, because of their condition, need replacing or major repair; or

2. Two or more other building components are old and, because of their condition, need replacement or major repair.

Key building components are those which, if in poor condition, could have an immediate impact on the integrity of the building and cause further deterioration in other components. They are the external components plus internal components that have potential safety implications and include:

- External walls
- Roof structure and covering
- Windows/doors
- Chimneys
- Central heating boilers
- Gas fires
- Storage heaters
- Plumbing
- Electrics

Other building components are those that have a less immediate impact on the integrity of the dwelling. Their combined effect is therefore considered, with a dwelling failing the disrepair standard if two or more elements are old and need replacing, or require immediate major repair.

Developing a Housing Stock Model for Wales

BRE have developed and used BRE Housing Stock Models for many years. These dwelling level models are used to estimate the likelihood of a particular dwelling meeting the criteria of interest. These outputs can then be mapped to provide a geographical distribution of each of the indicators. The process itself is made up of a variety of data sources, calculations and models.

The Housing Stock Model developed for Wales consists of the following datasets:

- OS AddressBase – this is an Ordnance Survey address list of all addresses. Addresses were extracted from it to form the residential address list for Wales
- OS Mastermap – this is the Ordnance Survey digital map which was used to inform dwelling type (i.e. detached house, terraced house, flat, etc.)
- Experian UK Consumer Dynamics Database of dwelling and household indicators – this was used to inform dwelling age, dwelling tenure and household composition
- Xoserve data – this is a database of mains gas availability at postcode level and was used to inform the likely fuel type of heating systems within dwellings

Other datasets were then integrated into this base model to provide enhanced information relating to dwelling tenure and dwelling energy data:

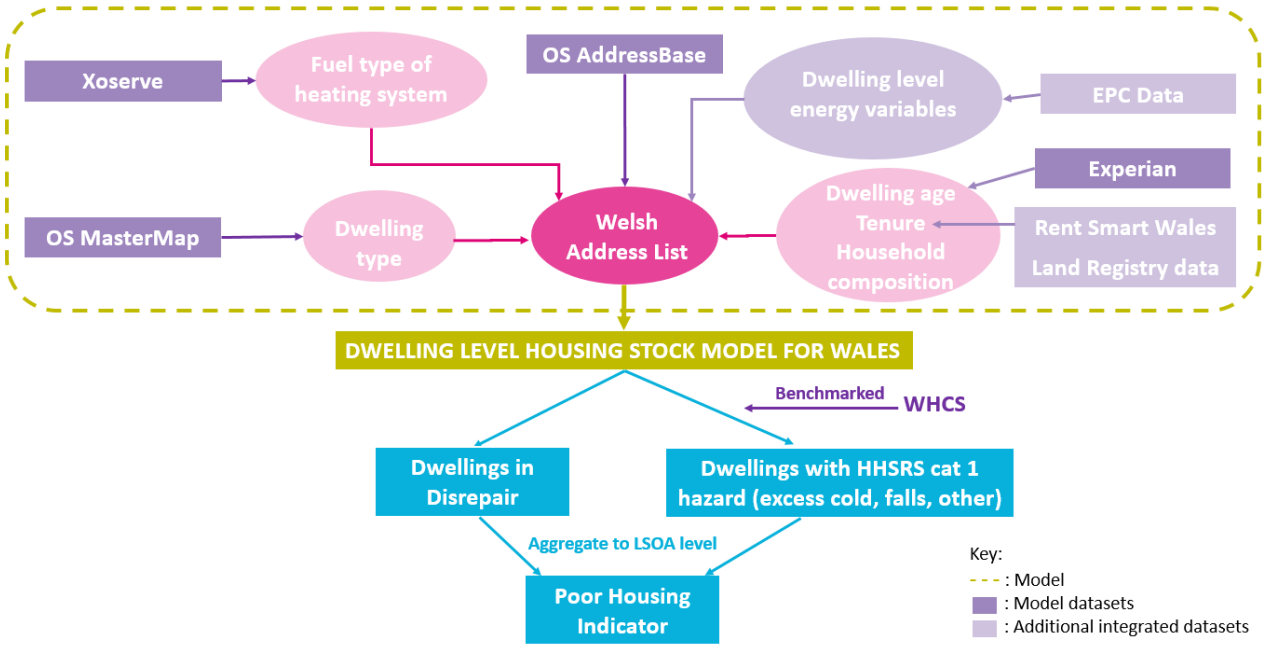
- RentSmart Wales data – this provided information relating to rented dwellings and was therefore used to inform dwelling tenure
- Land Registry Commercial and Corporate Ownership Data (CCOD) – this data was used to inform dwelling tenure by identifying social dwellings
- Energy Performance Certificate (EPC) data – this was used to inform energy variables within the Housing Stock Model

Model summary

The diagram below provides an overview of the BRE Housing Stock Model and the poor quality housing Indicator, showing how the different data sources listed above were used to develop or integrate within the model. The Welsh address list was used as the spine to link the other datasets and to form the base data for the model.

The dwelling level data for Wales was then used to produce the model for dwellings in disrepair and dwellings where at least one HHSRS Category 1 hazard was present.

Summary of the methodology to develop a Poor Housing Indicator for Wales



Determining HHSRS Category 1 hazards

BRE have developed a method for determining whether a dwelling fails the HHSRS Excess cold hazard: the BRE SimpleCO2 simplified energy model is designed to approximate SAP scores using a minimal level of data. This is an ideal tool for situations where dwellings have not been surveyed in detail but a measure of their energy efficiency is needed. For the remaining HHSRS hazards, BRE's standard logistic regression analysis was used.

Determining Disrepair

To determine dwellings in disrepair, logistic regression analysis was used to establish relationships between disrepair and various dwelling and social characteristics. Once these relationships were established, they were used to create a regression model which calculates the probability of a dwelling failing the disrepair criterion.

Determining the Poor Housing Indicator

Data from the Welsh House Condition Survey (WHCS) 2017-18 was used to benchmark the modelled estimates for disrepair and hazards, to Wales' totals. Once the benchmarked model based

on the Disrepair and HHSRS components had been created, they were combined to make an overall poor quality housing Indicator. If a dwelling failed one or more of the components, then it failed the poor quality housing Indicator. The dwelling level outputs for Wales were then aggregated to LSOA level.

Features of the model

BRE report that, for the disrepair component of the indicator, the main driver in the modelling process was dwelling age - the older the property the more likely it was to be in disrepair. There is also a tenure interaction with private rented properties increasing the likelihood of disrepair, followed by owner occupancy, then social rented properties being the least likely to be in disrepair. In terms of householder characteristics, the Experian mosaic classification of “senior security” is the main driver, with older financially secure households being less likely to occupy dwellings in disrepair.

In terms of the Category 1 hazards (HHSRS) component, the excess cold hazard was derived from BRE’s SimpleCO2 Model which estimated the energy efficiency of a dwelling. Areas with high levels of older dwellings off the gas network are typically where higher levels of excess cold are found. The main drivers for other hazards (falls and others) include dwelling age, again with older dwellings most likely to have Category 1 hazards. Dwelling type was also a factor, with flats less likely to have fall hazards. Tenure and household composition were also drivers, with private rented and larger households more likely to fail HHSRS.

Annex I: Flood Risk Category Weightings

The Flood Risk indicator considers the proportion of households at risk of flooding from rivers, the sea and surface water flooding. The risk is based on predicted frequency rather than the level of damage caused by flooding. The numbers of households at significant risk of flooding are given higher weighting than those at lower risk.

Due to a change in flooding categories, new weights were derived for WIMD 2014. The weighting methodology outlined in this annex was developed in partnership with Natural Resources Wales.

The way Flood Risk was communicated changed during 2013 and it was agreed that WIMD 2014 would use the new externally published NaFRA risk categories. New weighting factors were required for the new risk categories.

NaFRA Category	Definition (chance of flooding in any given year)	% of residential properties at risk of flooding	Weighting
High	Greater than or equal to 1 in 30	0.6%	0.06 (or 24)
Medium	Less than 1 in 30 but greater than or equal to 1 in 100	1.3%	0.01 (or 4)
Low	Less than 1 in 100 but greater than or equal to 1 in 1000	5.6%	0.0025 (or 1)
Very Low	Less than 1 in 1000.	0.05%	0.0025 (or 1)

In WIMD 2019, the Flood Risk indicator is produced from Flooding Risk Assessment Wales (FRAW) data provided by Natural Resources Wales. The FRAW data uses the same risk categories as the NaFRA data (with the exception of the 'very low' risk category) and therefore we have used the same flood risk weightings that were derived for WIMD 2014.

Methodology for determining weights

For WIMD 2014, it was proposed that we use information on average flooding damages as produced by the Middlesex University Flood Hazard Research Centre in their Multi Colour Manual. It was assumed that implications on quality of life increase as potential damage increases.

The damage values are assigned based on a sample of damages calculated to individual properties experiencing flooding from a range of modelled events for specific return-periods. No differentiation is made between the types of residential property.

Flood Frequency	Damage (£)
5	9,500
10	17,847
25	19,716
50	27,776
100	30,877

Source: Economic Appraisal Manual, 2013

The weighted annual average flood damage (WAAD) is the area under the graph of flood damages plotted against exceedance probability (the reciprocal of the return period in years). This is the area under the curve in Figure 1 below: Error! Reference source not found.**Figure 1 Damage vs Probability**

For properties in the high risk band, the probability of onset of flooding is 0.5. For properties in the medium and lower risk bands, the probability of the onset of flooding is reduced. The WAAD for each band is approximated to the area under each curve.

Assumptions

1. The weighted annual average damage calculation is based on properties within the extent of floodzone 2 of the NRW floodmap. This equates to properties with a greater than 0.1% annual chance of flooding.
2. For the UK in general, for increasingly extreme events the numbers of properties affected increases but average depths tend to be less than 1 m.
3. The very low flooding category is an artefact of combining NaFRA 50 m model squares with the extreme flood outline and affects a very small proportion of properties. Properties will be given the same low weighting.

Summary

Return Period	Exceedence Probability	WAAD Value	WAAD Ratio	WIMD Value	WIMD Ratio
30	0.0333	4976	23.5	0.0600	24
100	0.0100	767	3.6	0.0100	4
1000	0.0010	211	1	0.0025	1

Supporting data and calculations

Table 1 Source Table 4.32, "FCERM: Manual for Economic Appraisal 2013" FHRC

Return Period	Exceedence Probability	Damage (£)	Probability Interval	Mean Damage (£)	Annual Interval Damages (£)
2	0.500	0			
			0.300	4,750	1,425
5	0.200	9,500			
			0.100	13,674	1,367
10	0.100	17,847			
			0.060	18,782	1,127
25	0.040	19,716			
			0.020	23,746	475
50	0.020	27,776			
			0.010	29,327	293
100	0.010	30,877			
			0.005	30,877	154
200	0.005	30,877			
				WAAD	**4,842**

Table 2 High Band WAAD Calculation

Return Period	Exceedence Probability	Damage (£)	Probability Interval	Mean Damage (£)	Annual Interval Damages (£)
2	0.500	0			
			0.300	4,750	1,425
5	0.200	9,500			
			0.100	13,674	1,367
10	0.100	17,847			
			0.060	18,782	1,127
25	0.040	19,716			
			0.007	21,059	140
30	**0.033**	**22,403**			
			0.013	25,089	335
50	0.020	27,776			
			0.007	28,810	192
75	**0.013**	**29,843**			
			0.003	30,360	101
100	0.010	30,877			
			0.005	30,877	154
200	0.005	30,877			
			0.004	30,877	134
1,000	0.001	30,877			
				WAAD	**4,976**

Table 3 Medium Band WAAD Calculation

Return Period	Exceedence Probability	Damage (£)	Probability Interval	Mean Damage (£)	Annual Interval Damages (£)
2	0.500	0			
			0.300	0	0
5	0.200	0			
			0.100	0	0
10	0.100	0			
			0.060	0	0
25	0.040	0			
			0.007	0	0
30	**0.033**	**0**			
			0.013	13,888	185
50	0.020	27,776			
			0.007	28,810	192
75	**0.013**	**29,843**			
			0.003	30,360	101
100	0.010	30,877			
			0.005	30,877	154
200	0.005	30,877			
			0.004	30,877	134
1,000	0.001	30,877			
				WAAD	**767**

Table 4 Low Band WAAD Calculation

Return Period	Exceedence Probability	Damage (£)	Probability Interval	Mean Damage (£)	Annual Interval Damages (£)
2	0.500	0			
			0.300	0	0
5	0.200	0			
			0.100	0	0
10	0.100	0			
			0.060	0	0
25	0.040	0			
			0.007	0	0
30	**0.033**	**0**			
			0.013	0	0
50	0.020	0			
			0.007	0	0
75	**0.013**	**0**			
			0.003	0	0
100	0.010	0			
			0.005	15,439	77
200	0.005	30,877			
			0.004	30,877	134
1,000	0.001	30,877			
				WAAD	**211**

Annex J: Proximity to Accessible, Natural Green Space Calculation

Definition of accessible, natural green space

The definition for accessible, natural green space closely follows Natural Resource Wales' (NRW) Green Infrastructure Standards for Wales which are due for publication in December 2019.

The data set compiled to calculate the Proximity to Accessible, Natural Green Space indicator in WIMD 2019 comprises, and builds on, a series of rules for accessibility and naturalness that can confidently be said to be natural-feeling places to which the public have a legal right of access.

NRW and Welsh Government recognise that this definition is not all encompassing and that citizens may perceive many more spaces (including urban parks) as natural. Further, there are many more accessible spaces than are defined by law. Therefore, to enhance this definition, urban and coastal coverage – polygons representing urban and coastal green and blue spaces - have been used to supplement data gaps. Note that these additional polygons are 'likely' to be accessible but do not strictly indicate legal right of access.

This definition serves to show all land and water in Wales that are not covered in man-made surfaces and could therefore potentially deliver health or well-being benefits.

We are aware that this data set excludes some polygons which deliver ecosystem services (e.g. cycle paths, man-made play areas), but because the focus of this data set is on spaces which deliver health and well-being benefits by virtue of the natural nature of their surfaces we have deliberately excluded these.

NRW tested this rule-base by undertaking comparisons with aerial imagery as well as creating maps of familiar urban and rural areas to compare and verify the rule base delivered the intended result.

Data sources

Definition of accessible, natural greenspaces:

- All polygons from Ordnance Survey MasterMap Topography Layer® that satisfy the 'natural', 'multiple' and 'unknown' make classification. For these polygons to pass through to the final inclusion stage they must geographically intersect with;
- All polygons detailing Open Access from the Countryside Rights of Way Act 2000. The Act gives a right of access on foot for the purpose of open-air recreation and, in Wales, the right given under the Act commenced in May 2005. Layers included are:
    - Dedicated Forests
    - Other Dedicated Land
    - Open Country
    - Other Statutory Access Land
    - Registered Common Land

Supplementary green spaces to account for urban and coastal accessibility follow the definition of:

- All polygons from Ordnance Survey MasterMap Topography Layer® whose descriptive group satisfies the 'tidal water' classification.
- All 'natural', 'multiple' or 'unknown' polygons from Ordnance Survey MasterMap Topography Layer® that intersect within the extent of OS Open Greenspace polygons, that adhere to the following typologies:
    - Public parks or gardens
    - Sports areas or playing fields
    - Play space
    - Other sports facility

Routing methodology

Both network analysis and Euclidean distances (an 'as the crow flies' approach) were considered for modelling the travel-time for a residential dwelling's degree of proximity to accessible natural green spaces. However, whilst network analysis is highly representative of real-world behaviours and the favoured approach, Welsh Government recognise inconsistencies and significant data gaps in both greenspace polygon access point data and a nationally consistent path network. Therefore, applying network analysis would be nationally inconsistent and unfit for WIMD at this present time.

Although Euclidean distances do not take into account real-world obstacles, their production can be nationally consistent and provide a statistically robust method to calculate the areas of accessibility around greenspaces. Therefore Euclidean distances were used to produce the travel-time polygons around the green space data set for the purposes of WIMD 2019.

It is widely accepted that residential dwellings are classified as having sufficient access to accessible natural greenspaces if they can reach sites within a 5-minute walk. In a study undertaken by the University of Manchester for the formerly known Countryside Council of Wales, (now NRW) there is clear evidence to show that, generally, citizens are highly unlikely to walk beyond the 5-minute threshold to access local green infrastructure.

On average, a 5-minute walk roughly covers approximately 400 metres in distance. To account for the shortcomings of a Euclidean distance methodology, time-travel polygons were produced with a 300-metre radius to account for real-world obstacles in green space access.

All residential dwellings (as sourced from AddressBase® Plus) were then intersected with time-travel polygons to flag whether they were within a 5-minute walk to an accessible natural greenspace according to the definition.

Dwellings were then aggregated to Lower Super Output Area (LSOA) and the percentage of total dwellings within proximity of an accessible natural greenspace calculated.