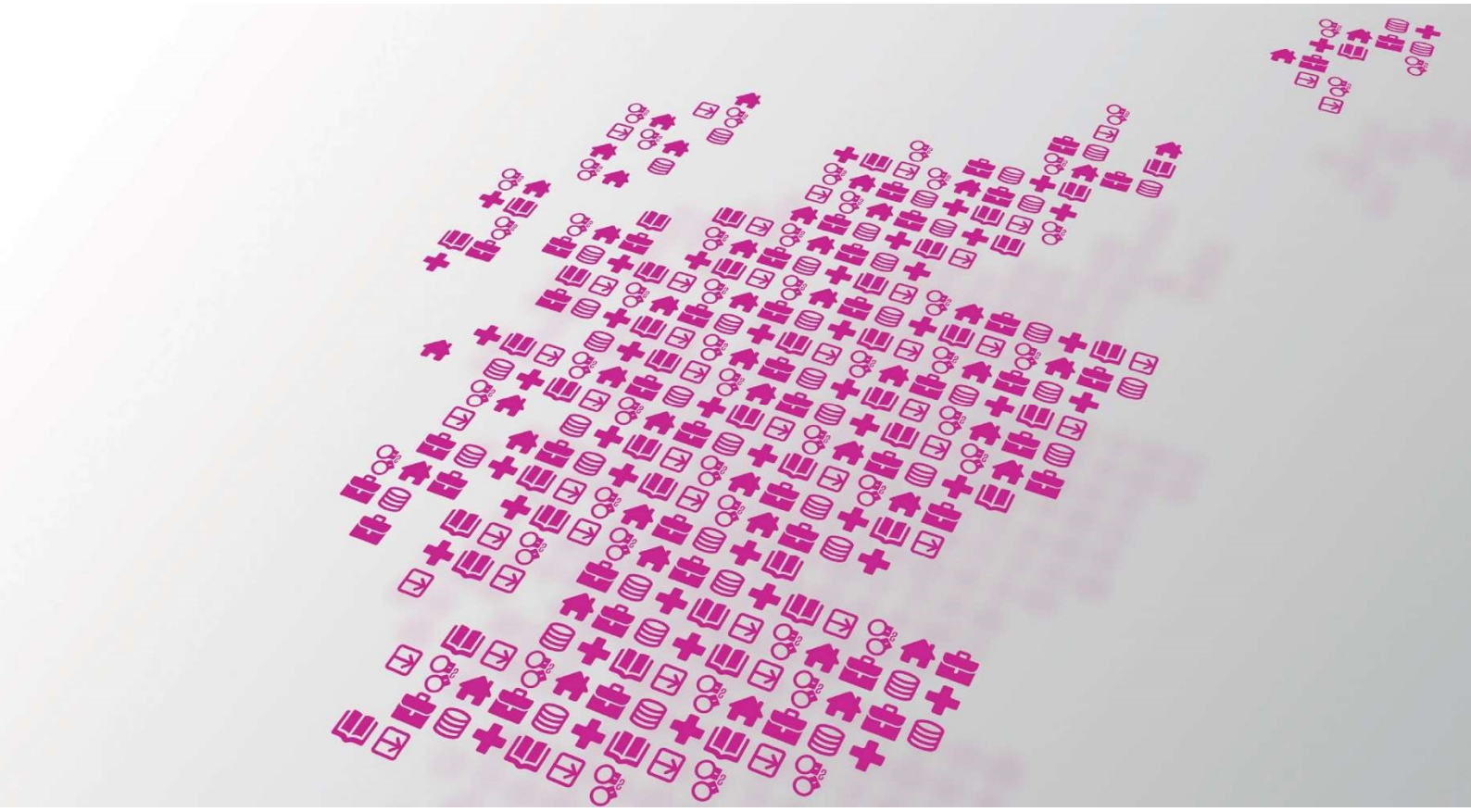
Scottish Index of Multiple Deprivation



SIMD 2020 technical notes

Contents

Introduction 4

Methodology overview 4

Constructing SIMD 5

- Constructing the domains 5
- Calculating SIMD 6
- Data providers 6

Changes to SIMD 2020 9

- Indicator changes 9

SIMD 2020 quality assurance 10

- Quality assurance process 10

Domains and indicators 12

Income domain 12

- Notes on the construction of the income domain 13
- Number of adults receiving Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance 14
- Number of adults receiving Guaranteed Pension Credit 15
- Number of dependent children (aged 0-18) for claimants of Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance 16
- People claiming Universal Credit and their dependent children (excluding those in the 'working with no requirements' conditionality group) 17
- Number of adults and children in Tax Credit families on low incomes 18

Employment domain 19

- Notes on the construction of the employment domain 19
- Working age recipients of Jobseeker's Allowance 20
- Working age recipients of Incapacity Benefit, Employment and Support Allowance or Severe Disablement Allowance 21
- Working age Universal Credit claimants not in employment 22

Health domain 23

- Notes on the construction of the health domain 24
- Standardised mortality ratio 25
- Hospital stays related to alcohol use: standardised ratio 27
- Hospital stays related to drug use: standardised ratio 29

Comparative illness factor: standardised ratio 31
Emergency stays in hospital: standardised ratio 33
Proportion of population being prescribed drugs for anxiety, depression or psychosis 35
Proportion of live singleton births of low birth weight 37
Education, skills and training domain 38
Notes on the construction of the education domain 38
School pupil attendance 40
Attainment of school leavers 42
Working age adults with no qualifications: standardised ratio 44
People aged 17-21 year olds entering university 46
People aged 16-19 not participating in education, employment or training 48
Geographic access to services domain 50
Notes on the construction of the access domain 50
Drive time sub-domain 51
Public transport sub-domain 54
Broadband sub-domain 56
Crime domain 58
Notes on the construction of the crime domain 58
Recorded crimes 60
SIMD crime codes 62
Housing domain 64
Notes on the construction of the housing domain 64
Persons in households that are overcrowded 65
Persons in households without central heating 67
ANNEX A 69
Department for Work and Pensions data geo-referencing 69
ANNEX B 70
Geographic access domain methodology 70

Introduction

The Scottish Index of Multiple Deprivation 2020 is the Scottish Government's standard tool for identifying concentrations of deprivation in Scotland. SIMD is based on work conducted by Oxford University in 1999. SIMD 2020 is the Scottish Government's sixth edition since 2004.

The SIMD 2020 technical notes provide information to assist researchers and analysts when using SIMD 2020 data (available online at www.gov.scot/SIMD). This document is intended to help users understand how SIMD is constructed, and as a reference when conducting further research or analysis based on SIMD and the associated data.

The following information is provided in this document:

- Methodology – the methods used to identify deprivation in Scotland and to calculate SIMD, along with information on how the data have been quality assured.
- Indicator details – key facts and background information on each of the indicators used in SIMD.

SIMD is one of four deprivation indices that cover the whole of the UK. The Scottish index differs from the other indices in the UK by following a slightly different methodology in constructing the overall index. The following are the key differences:

- SIMD is based on data zones which are smaller geographical units compared to the lower super output areas used in the other indices.
- The data sources are different for health, education, crime and housing domains.
- SIMD includes a specific domain on geographic access to services.

Methodology overview

SIMD combines seven different domains, or aspects, of deprivation. These are:

- Income
- Employment
- Health
- Education, skills and training
- Geographic access to services
- Crime
- Housing

These domains are constructed using a number of indicators to form ranks for each domain. Each of the seven domain ranks are then combined to form the overall SIMD. There are 6,976 data zones in Scotland. Data zones are ranked from 1 being most deprived to 6,976 being least deprived. This provides a measure of relative deprivation at data zone level, therefore it tells you that one data zone is relatively more deprived than another, but not how much more deprived.

The methodology used to construct SIMD 2020 remains fundamentally the same as that used to construct the previous versions of SIMD in 2004, 2006, 2009, 2012 and

2016. It is based on the methodology developed by Oxford University to produce the Scottish Indices of Deprivation in 2003[1]. The Scottish Government produced the first SIMD in-house in 2004. As the methodology used has remained broadly consistent in each subsequent update, these technical notes provide a summary of how the SIMD is constructed and details on the indicators included in each domain of SIMD 2020. Full details of the individual methods for creating the domains and overall index are described in the SIMD 2004 Technical Report[2].

Constructing SIMD

SIMD 2020 is built up from a total of 33 indicators covering the seven domains. A list of the indicators included in each domain is provided in the following domain chapters. The indicators for each domain were selected on the basis that they are:

- domain-specific and appropriate for the purpose (as direct as possible measures for the given type of deprivation);
- up-to-date;
- capable of being updated on a regular basis;
- statistically robust;
- measure major features of a given type of deprivation (not conditions just experienced by a very small number of people or areas).

The domains are calculated differently depending on the type of data used in each one. This is explained in more detail below and illustrated visually on page **Error! Bookmark not defined.**.

Constructing the domains

The income, employment and housing domains are created by summing counts of people and dividing by the appropriate population denominator (taken form the NRS Small Area Population Estimates (SAPE) or Census). The crime domain and some indicators in the health and education domain also use SAPE.

The **income domain** is constructed by adding all five income indicators and dividing by the 2017 mid-year total population from SAPE. Thus, the domain score is a simple percentage.

The **employment domain** is constructed by adding the three employment indicators and dividing by the 2017 mid-year working age population estimates taken from SAPE. The domain score is a simple percentage.

The **health, education and access domains** are created by ranking the indicators and transforming them to a standard normal distribution. This standardisation process is necessary because the indicators in these domains may be measured in different ways and on different scales. A statistical technique called factor analysis is

[1] https://www2.gov.scot/Topics/Statistics/scotstat/snsgroup/InterimIndices2003

[2] https://www.webarchive.org.uk/wayback/archive/20180516204453mp_/http://www.gov.scot/Resource/Doc/26350/0025281.pdf

then used to create a weight for each indicator. Next, the indicators are combined to produce a domain score which is then ranked.

The **crime domain** is a count of selected recorded crimes, divided by the 2017 mid-year population estimates from SAPE. It is presented as the crime rate per 10,000 population.

The **housing domain** is the sum of the two housing indicators, divided by the total household population from the 2011 Census. The domain score is a simple percentage.

Calculating SIMD

Once the individual domain scores are calculated, they are combined to create the overall SIMD. The overall SIMD is a weighted sum of the seven domain scores, with different domains given different weights. The weighting is based on the original research conducted by Oxford University, when the original Scottish Index of Multiple Deprivation was first produced, and takes into consideration how up to date and robust the indicators within each domain are.

The domain weighting used for SIMD 2020 remains the same as in SIMD16. A review of the weighting was undertaken when preparing for SIMD16 and concluded that the changes to data quality and methodology were not enough to justify a change of weightings.

The table below shows the percentage of overall SIMD 2020 for each of the seven domains:

Domain	Percentage of overall SIMD 2020
Income	28%
Employment	28%
Health	14%
Education, skills and training	14%
Geographic access to services	9%
Crime	5%
Housing	2%

Prior to the weighting, the domains are standardised by ranking the scores. The ranks then undergo exponential transformation to avoid high ranks in one domain cancelling out low ranks in another. The resulting scores for the overall SIMD are then ranked from 1 (most deprived) to 6,976 (least deprived) to create the final index.

The SIMD 2020 methodology is shown in the flow diagram below.

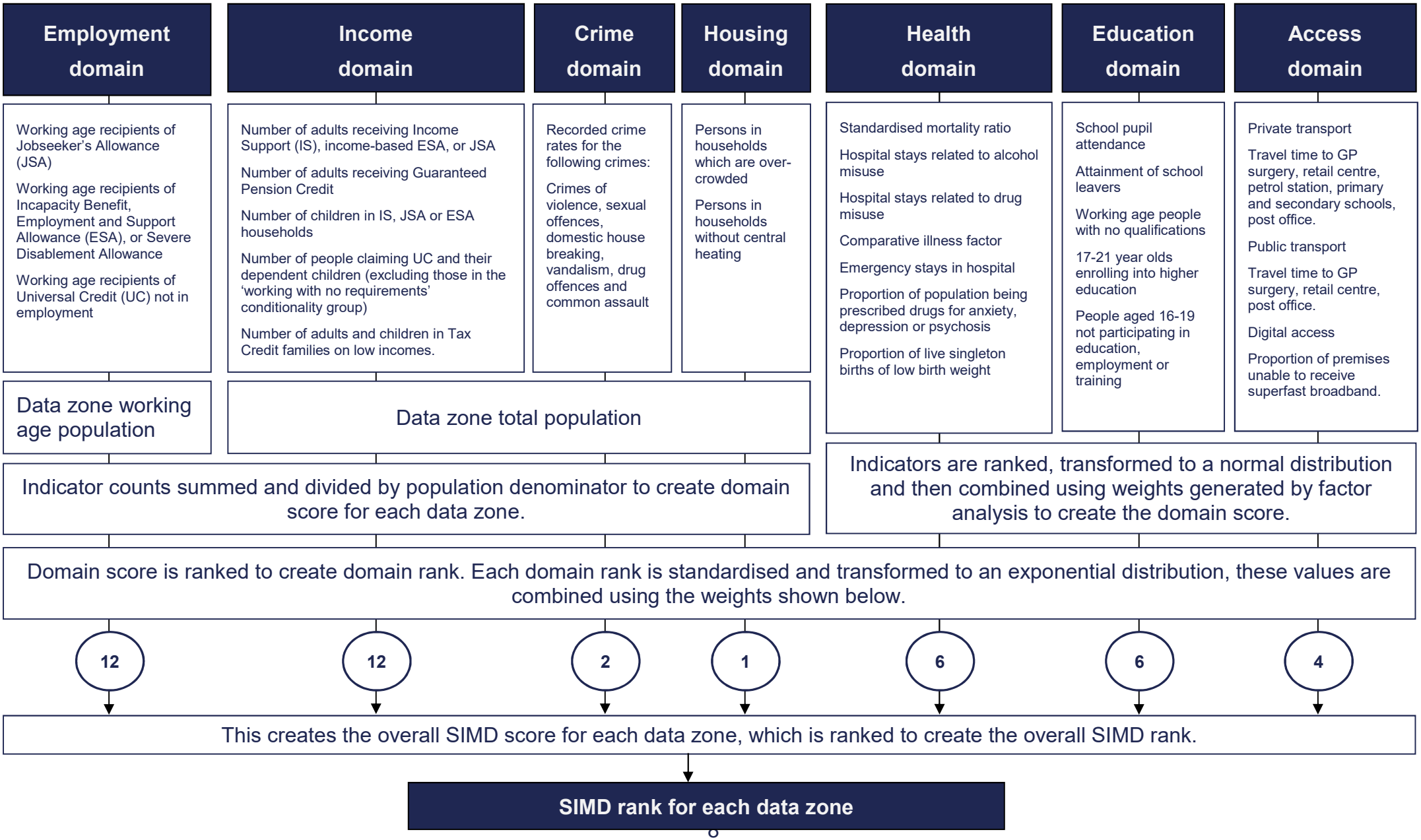
Data providers

The following external organisations provided data for the construction of SIMD 2020:

- Department for Work and Pensions
- Her Majesty's Revenue and Customs

- National Records of Scotland
- NHS Scotland Information Services Division
- Scottish Qualifications Authority
- Higher Education Statistics Agency
- Skills Development Scotland
- Police Scotland
- Ofcom

SIMD 2020 methodology



Changes to SIMD 2020

Indicator changes

SIMD 2020 is broadly consistent with the previous SIMD16 publication. Changes to SIMD 2020 include improvements to the education and access domains, along with changes to a number of indicators to take account of the introduction of Universal Credit. Further substantial development work will be required for the next SIMD release, potentially after the publication of the next Census. The box below summarises the main changes made to SIMD 2020 within each domain. Full details of the indicators included in each domain are provided in the following chapter.

Summary of changes to SIMD 2020 domains and indicators

Income domain

There have been minor changes to some indicators to ensure consistency across the different measures and data sources.

The number of adults not in paid employment receiving UC has been replaced with the number of people claiming UC and their dependent children aged 0-18 (excluding those in the 'working with no requirements' conditionality group). These changes were informed by work carried out by DWP to assess the best way of capturing people who would have received different benefits before the roll-out of UC.

Employment domain

Changes to the welfare system also affected the employment domain. The ONS unemployment claimant count was discontinued and been replaced by working age recipients of Jobseeker's Allowance.

A new indicator for working age Universal Credit claimants not in employment has also been introduced for SIMD 2020. This decision was informed by work carried out by DWP to assess the best way of capturing people who would have received different benefits before the roll-out of Universal Credit.

Health domain

There has been a change in the definition of drug-related conditions, which is used to define hospital stays related to drug use. Following a consultation with users of statistics, the list of drugs for identifying drug-related conditions was expanded by including a number of T-codes from the International Classification of Diseases 10th revision (ICD-10).

The comparative illness factor measure reflects changes in the welfare system, including the introduction of Universal Credit.

Education domain

The DWP measure of people aged 16-19 not in full time education, employment or training (NEET) has been replaced by a new indicator based on the annual participation measure dataset provided by Skills Development Scotland.

Access to services domain

The software and data sources used to calculate journey times has changed. Details can be found in Annex B.The domain now includes an access to digital services indicator.

SIMD 2020 quality assurance

Data used for SIMD indicators are created from administrative systems and come from different data providers. Administrative data are collected by government departments and other organisations and consist of information gathered primarily for operational reasons, such as registration, transactions and record-keeping, usually when delivering a service. Their statistical use is therefore secondary.

As this type of data are not collected for analysis and research purposes, there are limitations with their use when compiling official statistics. The SIMD team is proactive in quality assuring and investigating the data to ensure they are fit for purpose.

Quality assurance process

The quality assurance process for SIMD 2020 consisted of checks at various stages of the process:

- indicator QA conducted by data providers and the SIMD team
- indicator QA conducted by local authority analysts
- domain QA conducted by external partners with experience of the domain subject.

The section below provides more information on these stages and the parties involved in each stage of the process.

Indicator QA - data providers: in the initial stage of the QA process data providers were expected to carry out appropriate checks to ensure data were complete and in the correct formats for calculating the indicators. Data providers were required to inform the SIMD team of any issues in the datasets they were aware of that could affect the quality of SIMD 2020. They were also asked to complete a technical report on the data and any analysis they provided.

Indicator QA - SIMD team: after receiving the indicator data from the data providers, the SIMD team conducted its initial QA checks. As data zone boundaries remained the same as for SIMD16, the QA process was mostly based on comparing datasets for SIMD 2020 with datasets for SIMD16. Outliers were therefore identified by comparing indicator distributions, range, minimum and maximum values. Any notable differences that could not be easily explained were recorded for further investigations.

Indicator QA - local authorities: the SIMD team invited local authorities to get involved in the indicator QA for their council area and provided them with the outliers identified for each indicator. Local authority analysts were asked to sense-check the data, using their local knowledge and experience to justify, where possible, conspicuous variations. Local authorities were asked to complete QA forms and return them to the SIMD team. Where consistent patterns of changes or considerable variations were observed, the SIMD team reported them to data providers for further investigation.

The access domain was quality assured separately due to its technical complexity.

Indicators for the housing domain did not require checking as the data used for SIMD 2020 are based on the 2011 Census.

Domain QA - external partners: after the comments were addressed, the SIMD team calculated the SIMD domains. Domain ranks were then sent to external partners with relevant experience on the different domain subjects for their quality checks. The domain QA consisted in comparing the new domain ranks with SIMD16 domain ranks and detecting and large or unexpected movements of numbers of data zones between each SIMD decile.

SIMD QA - SIMD team: some additional checks were carried out by the SIMD team after calculating the overall index to ensure the calculation steps were correct. These included looking at changes in the relative position of each data zone in the overall index, movements of numbers of data zones between each decile, changes in local and national shares for each local authority and comparing boxplots for each local authority.

The SIMD team would like to thank all data providers, local authority and health board analysts, and other internal and external colleagues, for their help in quality assuring SIMD 2020 and for providing meaningful insight and knowledge for interpreting the data.

A particular note of thanks is due to Fliss Rollings from the Scottish Government Geographic Information Science and Analysis Team and Dr David Connolly, Director of Innovation at SYSTRA Ltd, for respectively constructing and quality assuring the access domain, and Nick Bailey and Luis Serra from University of Glasgow for quality assuring the employment and income domains.

Domains and indicators

Income domain

SIMD 2020 indicators	SIMD16 indicators	Summary of change
Number of adults receiving Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance Source: DWP, August 2017	Number of adults (aged 16-59) receiving Income Support (IS) or income-based Employment and Support Allowance (ESA), and the number of adults (all ages) receiving Jobseeker's Allowance (JSA) Source: DWP, August 2015	Eligibility criteria for certain benefits have changed and Universal Credit was introduced.
Number of adults receiving Guaranteed Pension Credit DWP, August 2017	Number of adults (aged 60 plus) receiving Guaranteed Pension Credit DWP, August 2015	No change
Number of children (aged 0-18) dependent on a recipient of Income Support, Jobseeker's Allowance or Employment and Support Allowance DWP, August 2017	Number of children (aged 0-15) dependent on a recipient of Income Support, Jobseeker's Allowance or Employment and Support Allowance DWP, May 2015	Eligibility criteria for certain benefits have changed and Universal Credit was introduced.
Number of people claiming Universal Credit and their dependent children (aged 0-18) (excluding those in the 'working with no requirements' conditionality group) DWP, August 2017	Number of adults (all) not in paid employment receiving Universal Credit DWP, August 2015	Changes to this indicator were informed by work carried out by DWP to assess the best way of capturing people who would have received different benefits before the roll-out of Universal Credit.
Number of adults and children in Tax Credit families on low incomes HMRC, August 2017	Number of adults and children in Tax Credit families on low incomes HMRC, 2013-14	No change

Notes on the construction of the income domain

The income domain is calculated by adding the component counts and dividing by the 2017 mid-year population estimate (source: NRS).

Missing data

In some data zones, the resident population was zero in the year considered, and income rates could not be determined. These data zones were excluded from the ranking calculation.

The following tables give more detailed information on the indicators in the income domain. Due to disclosure issues, SIMD 2020 datasets only include the combined counts for the whole domain.

Number of adults receiving Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance

General description of indicator	The number of adults receiving Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance.
Indicator type	Count
Time period	August 2017
Data source	DWP
Method of construction of indicator	The data were extracted from the Work and Pensions Longitudinal Study and the domain was constructed by DWP.
Key decisions on methodology	There are no methodological changes to this indicator compared to SIMD16. However, there have been changes to the welfare system, such as the roll-out of Universal Credit.
Comparison with 2016 indicator	This indicator cannot be compared directly with earlier time periods due to changes in the welfare system, which affect the eligibility criteria for certain benefits.
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html.
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the income domain are not available. Instead, a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Number of adults receiving Guaranteed Pension Credit

General description of indicator	The number of adults receiving Guarantee Pension Credit.
Indicator type	Count
Time period	August 2017
Data source	DWP
Method of construction of indicator	The data were extracted from the Work and Pensions Longitudinal Study and the domain was constructed by DWP.
Key decisions on methodology	Indicator remains the same as SIMD 2016.
Comparison with 2016 indicator	This indicator cannot be compared directly with earlier time periods due to changes in the welfare system, which affect the eligibility criteria for certain benefits.
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html.
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the income domain are not available. Instead, a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Number of dependent children (aged 0-18) for claimants of Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance

General description of indicator	The number of dependent children (aged 0 -18) of claimants of Income Support, income-based Employment and Support Allowance, or Jobseeker's Allowance.
Indicator type	Count
Time period	August 2017
Data source	DWP
Method of construction of indicator	Claimants were taken from the 100% Sure Start dataset, and dependent children were matched to claimants from the Child Benefit Scan.
Key decisions on methodology	There are no methodological changes to this indicator compared to SIMD16. However, there have been changes to the welfare system, such as the roll-out of Universal Credit.
Comparison with 2016 indicator	This indicator cannot be compared directly with earlier time periods due to changes in the welfare system, which affect the eligibility criteria for certain benefits.
Implications of comparing this indicator with SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html.
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the income domain are not available. Instead, a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

People claiming Universal Credit and their dependent children (excluding those in the 'working with no requirements' conditionality group)

General description of indicator	All Universal Credit (UC) claimants (excluding those in the 'working with no requirements' conditionality group) and their dependent children aged 18 and under.
Indicator type	Count
Time period	August 2017
Data source	DWP
Method of construction of indicator	The indicator was constructed using the 'children in out-of-work benefit households' dataset merged with the UC caseload dataset.
Key decisions on methodology	This indicator captures people who would have previously received legacy benefits and have now come under the UC regime. Changes to this indicator were informed by work carried out by DWP and Welsh Government analysts to assess the best way of capturing people who would have received a different set of benefits before the roll-out of Universal Credit.
Comparison with 2016 indicator	This is a new indicator.
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html.
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the income domain are not available. Instead, a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Number of adults and children in Tax Credit families on low incomes

General description of indicator	The number of adults and children (aged 0-15 and dependents aged 16-19) in in-work families claiming Working or Child Tax Credit with an income of less than the poverty threshold (equivalised income less than £222.55 per week). Out-of-work claimants of Working or Child Tax Credit not getting Income Support or Jobseeker's Allowance, and dependent children (as above).
Indicator type	Count
Time period	August 2017
Data source	HM Revenue and Customs
Method of construction of indicator	The data were extracted from the HMRC Tax Credit system and supplied to DWP where the domain was constructed.
Key decisions on methodology	Only those households with an equivalised weekly income of less than 60% of the median (£222.55) were included in the indicator, as this cut-off is used to identify those families on low incomes.
Comparison with 2016 indicator	
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html.
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the income domain are not available. Instead, a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Employment domain

SIMD 2020 indicators	SIMD16 indicators	Summary of change
Working age recipients of Jobseeker's Allowance Source: DWP, Feb, May, Aug, Nov 2017 averaged	Working age unemployment claimant count averaged over 12 months ONS, 2014	ONS Claimant Count discontinued.
Working age recipients of Incapacity Benefit, Employment and Support Allowance, or Severe Disablement Allowance DWP, Feb, May, Aug, Nov 2017 averaged	Working age recipients of Incapacity Benefit, Employment and Support Allowance, or Severe Disablement Allowance DWP, August 2015	No change
Working age Universal Credit claimants not in employment DWP, Feb, May, Aug, Nov 2017 averaged	-	New indicator for SIMD 2020

Notes on the construction of the employment domain

The employment domain is calculated by adding the three sets of counts and dividing by the 2017 mid-year working age population estimate (source: NRS).

Missing data

In some data zones, the resident population was zero in the year considered and employment rates could not be determined. These data zones were excluded from the ranking calculation.

The following tables give more detailed information on each of the indicators in the employment domain used in constructing SIMD 2020. However, SIMD 2020 datasets only report the combined totals for the whole domain.

Working age recipients of Jobseeker's Allowance

General description of indicator	The number of working age recipients of Jobseeker's Allowance.
Indicator type	Count
Time period	Feb 2017, May 2017, August 2017 and Nov 2017 (averaged)
Data source	DWP Work and Pensions Longitudinal Study.
Method of construction of indicator	DWP provided averaged data for Feb, May, Aug and Nov 2017.
Key decisions on methodology	Due to the introduction of Universal Credit, the claimant count published by ONS is no longer considered to provide an accurate estimate of unemployment rates. This also affects JSA data to some extent, therefore the employment domain includes UC claimants who would otherwise been eligible for JSA.
Comparison with 2016 indicator	This indicator is not comparable with the SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the employment domain are not available, instead a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Working age recipients of Incapacity Benefit, Employment and Support Allowance or Severe Disablement Allowance

General description of indicator	The number of working age recipients of Incapacity Benefit, Employment and Support Allowance or Severe Disablement Allowance.
Indicator type	Count
Time period	Feb 2017, May 2017, August 2017 and Nov 2017 averaged.
Data source	DWP Work and Pensions Longitudinal Study.
Method of construction of indicator	The data were extracted from the WPLS and the domain was constructed by DWP.
Key decisions on methodology	This indicator remains the same as in the SIMD 2016. However, some people who may have received Employment and Support Allowance will now be covered by UC regime. The employment domain therefore includes relevant UC claimants to account for this.
Comparison with 2016 indicator	This indicator cannot be compared directly with earlier time periods due to changes in the welfare system, which affect the eligibility criteria for certain benefits.
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the employment domain are not available. Instead, a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Working age Universal Credit claimants not in employment

General description of indicator	The number of working age recipients of Universal Credit not in employment.
Indicator type	Count
Time period	Feb 2017, May 2017, August 2017 and Nov 2017 averaged
Data source	DWP UC caseload dataset.
Method of construction of indicator	The data were extracted from DWP UC caseload dataset.
Key decisions on methodology	This is a new indicator for SIMD 2020. Changes to this indicator were informed by work carried out by DWP and Welsh Government to assess the best way of capturing people who would have received a different set of benefits before the roll-out of Universal Credit.
Comparison with 2016 indicator	
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	The population estimate for a few data zones was zero in 2017, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	Random error in line with DWP policy was used. Further information on this approach can be found at https://stat-xplore.dwp.gov.uk/webapi/metadata/Data%20Confidentiality.html
Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The individual datasets used to construct the employment domain are not available, instead a combined count and a combined rate per population of all indicators is available at www.gov.scot/SIMD.

Health domain

SIMD 2020 indicators	2020 weight	SIMD16 indicators	2016 weight	Summary of change
Standardised mortality ratio Source: NRS, 2014-15 to 2017-18	0.06	Standardised mortality ratio ISD, 2011-14	0.07	No change
Hospital stays related to alcohol use: standardised ratio ISD, 2014-15 to 2017-18	0.08	Hospital stays related to alcohol use: standardised ratio ISD, 2011-14	0.10	No change
Hospital stays related to drug use: standardised ratio ISD, 2014-15 to 2017-18	0.07	Hospital stays related to drug use: standardised ratio ISD, 2011-14	0.06	Change in the definition of drug-related conditions
Comparative illness factor: standardised ratio DWP, Aug 2017	0.46	Comparative illness factor: standardised ratio DWP, Aug 2015	0.37	Changes in the welfare system.
Emergency stays in hospital: standardised ratio ISD, 2014-15 to 2017-18	0.19	Emergency stays in hospital: standardised ratio ISD, 2011-15	0.22	No change
Proportion of population being prescribed drugs for anxiety, depression or psychosis ISD, 2017-18	0.13	Proportion of population being prescribed drugs for anxiety, depression or psychosis ISD, 2014-15	0.15	No change
Proportion of live singleton births of low birth weight ISD, 2014-15 to 2017-18	0.01	Proportion of live singleton births of low birth weight ISD, 2011-14	0.02	No change

Notes on the construction of the health domain

Missing data

For some indicators, the data zone population was zero during the time period considered, and therefore indicator rates and ranks could not be determined. Missing rates and ranks are denoted by '*'.

To calculate the overall rankings for the health domain, the normalised scores for these data zones were set to zero before combining the indicators. As a result, the indicators with missing values moved the overall domain ranking of these data zones towards a middle ranking.

Weights

The weightings of the individual indicators in the health domain are rounded and may therefore not add up to 100 per cent.

Age-sex standardisation

Five of the seven indicators in the health domain are indirectly standardised ratios.

The aim of standardisation is to provide a summary 'adjusted' rate to take into account underlying differences (for example age, sex, deprivation) of a study population relative to a 'reference' population.

Indirect age-sex standardisation is based on a comparison of observed to expected numbers of events or cases, achieved by applying age-specific rates from a 'standard population' to the population of interest. For example, if the study population is within a data zone of residence then the standard population might be taken as Scotland.

For more information on direct and indirect standardisation techniques please refer to the NHS guidance[3].

The following tables give more detailed information on each of the indicators in the health domain used in constructing SIMD 2020.

[3] https://www.scotpho.org.uk/media/1400/phi-standardisation-guidance-v21.docx

Standardised mortality ratio

General description of indicator	Indirectly age-sex standardised ratio for deaths of all ages registered from all causes. Data standardised by 5-year age band and sex.
Indicator type	Indirectly standardised ratio: four-year period
Time period	Financial years 2014-15 to 2017-18
Data source	National Records of Scotland (NRS)
Denominator used	Expected events (calculated from indirect standardisation to Scotland)
Data source of denominator	NRS mid-year population estimates and mortality information
Method of construction of indicator	For each data zone, the sum of observed all-cause deaths by sex and 5-year age band was divided by the corresponding sum of expected deaths. For further details, please see methodology section (2.3) of SIMD 2006 technical report[4].
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	Indirectly standardised ratios have limitations for drawing comparisons. In comparing one data zone between different SIMD releases, if it had a ratio of 100 in SIMD16 and 110 in SIMD 2020, you could say that the figure for that data zone has got worse **relative to Scotland**. However, you cannot say whether the data zone has actually worsened – only that it has worsened relative to Scotland. In comparing two data zones for the same SIMD release (e.g. SIMD 2020), if one data zone had a ratio of 110 and another a ratio of 115, you could say that the figures for both data zones are worse than the Scottish average (100). However, for statistical reasons, you **cannot** say that the data zone with the ratio of 115 is worse than the one with the ratio of 110.

[4] https://www2.gov.scot/Resource/Doc/933/0041180.pdf

Other data quality issues	Given the legal requirements around death registrations, it is expected that these data do not have data quality issues.
Disclosure control	N/A
Geo-referencing	All postcodes are validated at source. Data exclude people where no match to a data zone was possible (e.g. homeless, incomplete postcode information).
Availability of data	The raw data for this indicator have not been published. The standardised ratio is available on www.gov.scot/SIMD.

Hospital stays related to alcohol use: standardised ratio

General description of indicator	Indirectly age-sex standardised ratio of observed to expected stays in acute NHS hospitals in Scotland with a diagnosis of alcohol-related conditions (based on any of six possible diagnoses), both sexes, all ages. These figures are based on the continuous inpatient stay (CIS). Individual SMR01 (acute hospital) episodes for each patient are linked together using probability matching to create "linked" patient histories and CISs (continuous stays in hospital regardless of whether or not this involves transfer between hospitals or between specialties within the same hospital).
Indicator type	Indirectly standardised ratio: four-year period
Time period	Financial years 2014-15 to 2017-18
Data source	NHS Scotland Information Services Division (ISD), Scottish Morbidity Record (SMR01)
Denominator used	Expected stays (calculated by indirect standardisation)
Data source of denominator	ISD SMR01 and NRS mid-year resident population estimates
Method of construction of indicator	For each data zone, the sum of observed CISs by sex and 5-year age band was divided by the corresponding sum of expected CISs. For further details, please see methodology section (2.3) of SIMD 2006 technical report[5]. Alcohol-related conditions are defined using the World Health Organization's International Classification of Diseases 10th revision (ICD-10) (E244, E512, F10, G312, G621, G721, I426, K70, K292, K852, K860, O354, P043, Q860, R780, T510, T511, T519, X45, X65, Y15, Y90, Y91, Y573, Z502, Z714, Z721).
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.

[5] https://www2.gov.scot/Resource/Doc/933/0041180.pdf

Implications of comparing this indicator with the one used in SIMD 2016	Indirectly standardised ratios have limitations for drawing comparisons. In comparing one data zone between different SIMD releases, if it had a ratio of 100 in SIMD16 and 110 in SIMD 2020, you could say that the data zone has got worse **relative to Scotland**. However, you cannot say whether the data zone has actually worsened – only that it has worsened relative to the overall Scotland level. In comparing two data zones for the same SIMD release, if one data zone had a ratio of 110 and another a ratio of 115, you could say that both data zones are worse than the Scottish average (100), but for statistical reasons you **cannot** say that the data zone with the ratio of 115 is worse than the one with the ratio of 110.
Other data quality issues	There were data zones with no alcohol-related CISs in the time-period used, and therefore the ratios and ranks are identical in these cases. Caution is necessary when interpreting these figures as recording may vary from hospital to hospital. Where an alcohol-related condition is suspected but unconfirmed, it may not be recorded.
Disclosure control	N/A
Geo-referencing	All postcodes are validated at source. Data exclude people where no match to a data zone was possible (e.g. homeless, incomplete postcode information).
Availability of data	The standardised ratio is available on www.gov.scot/SIMD.

Hospital stays related to drug use: standardised ratio

General description of indicator	Indirectly age-sex standardised ratio of observed to expected stays in acute NHS hospitals in Scotland with a diagnosis of drug use or poisoning, both sexes, all ages. Figures are based on continuous inpatient stays (CIS) from linked general acute hospital inpatient & day case (SMR01) episodes. Each ratio is expressed as a number. A ratio value of 100 means that the observed and expected number of stays for a data zone was exactly the same. A value lower than 100 indicates that there were fewer stays than expected. A value higher than 100 indicates that there were more stays than expected.
Indicator type	Indirectly standardised ratio: four-year period
Time period	Financial years 2014-15 to 2017-18
Data source	NHS Scotland Information Services Division (ISD), Scottish Morbidity Record (SMR01)
Denominator used	Expected stays (calculated by age-sex standardisation)
Data source of denominator	ISD SMR01 and NRS mid-year resident population estimates
Method of construction of indicator	For each data zone, the sum of observed CISs by sex and 5-year age band was divided by the corresponding sum of expected CISs. For further details, please see methodology section (2.3) of SIMD 2006 technical report[6]. Drug-related conditions are coded using the World Health Organization's International Classification of Diseases 10th revision (ICD-10). The following ICD-10 codes were used to identify drug-related conditions: F11-F16, F18, F19, T40.0, T40.1, T40.3 and T40.5-T40.9.
Key decisions on methodology	Following the consultation with users of statistics, the list of drugs for identifying drug-related conditions was expanded by including ICD-10 T-codes which are listed above. For SIMD16, only the ICD-10 F-codes were used (also listed above).
Comparison with 2016 indicator	Due to the change in the definition of drug-related conditions, this indicator for SIMD 2020 is not directly comparable with the indicator for SIMD16.

[6] https://www2.gov.scot/Resource/Doc/933/0041180.pdf

Implications of comparing this indicator with the one used in SIMD 2016	Indirectly standardised ratios have limitations for drawing comparisons. In comparing one data zone between different SIMD releases, if it had a ratio of 100 in SIMD16 and 110 in SIMD 2020, you could say that the data zone has got worse **relative to Scotland**. However, you cannot say whether the data zone has actually worsened – only that it has worsened relative to Scotland. In comparing two data zones for the same SIMD release, if one data zone had a ratio of 110 and another a ratio of 115, you could say that both data zones are worse than the Scottish average (100), but for statistical reasons you **cannot** say that the data zone with the ratio of 115 is worse than the one with the ratio of 110.
Other data quality issues	There were data zones with no drug-related CISs in the time-period used, and therefore the ratios and ranks are identical in these cases. Caution is necessary when interpreting the figures. Recording of drug use may vary from hospital to hospital. Where a drug-related condition is suspected but unconfirmed, it may not be recorded by the hospital. Further, where drug use is recorded, it may not be possible to identify which drugs were involved.
Disclosure control	N/A
Geo-referencing	All postcodes are validated at source. Data exclude people where no match to a data zone was possible (e.g. homeless, incomplete postcode information).
Availability of data	The standardised ratio is available on www.gov.scot/SIMD.

Comparative illness factor: standardised ratio

General description of indicator	The Comparative Illness Factor (CIF) is a combined count of the total number of people claiming one or more of: Disability Living Allowance, Attendance Allowance, Incapacity Benefit (not receiving DLA), Employment and Support Allowance, Severe Disablement Allowance, Income Support with disability premium, Personal Independence Payment and Universal Credit claimants with an accepted restricted ability to work.
Indicator type	Indirectly standardised ratio
Time period	August 2017
Data source	DWP, Work and Pensions Longitudinal Study and Universal Credit dataset.
Denominator used	Expected frequency of claiming the selected benefits, calculated during age-sex standardisation (see SIMD 2006 technical report[7] for details).
Data source of denominator	NRS Small Area Population Estimates 2017, DWP and WPLS
Method of construction of indicator	The observed number of benefit recipients was divided by the expected number for each data zone to provide a ratio. For further details of the standardisation method see the SIMD 2006 technical report.
Key decisions on methodology	This indicator reflects changes in the welfare system including the introduction of Universal Credit.
Comparison with 2016 indicator	
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	After calculating the standardisation there are cases where division by zero occurs due to population estimate of data zone being zero. Such cases are marked "*".
Disclosure control	Figures are rounded to the nearest 5. All benefit figures for each geography breakdown have also had a 'blurring' factor applied in line with the DWP data disclosure policy.

[7] http://www.scotland.gov.uk./Publications/2006/10/13142913/0

Geo-referencing	See Annex A for explanation of geo-referencing of DWP data.
Availability of data	The standardised ratio is available on www.gov.scot/SIMD.

Emergency stays in hospital: standardised ratio

General description of indicator	Indirectly age-sex standardised ratio of observed to expected emergency stays in acute NHS hospitals in Scotland, both sexes and all ages. These figures are based on the continuous inpatient stay (CIS). Individual SMR01 (acute hospital) episodes for each patient are linked together using probability matching to create "linked" patient histories and CISs (continuous stays in hospital regardless of whether or not this involves transfer between hospitals or between specialties within the same hospital).
Indicator type	Indirectly standardised ratio: four-year period
Time period	Financial years 2014-15 to 2017-18
Data source	NHS Scotland Information Services Division (ISD), Scottish Morbidity Record (SMR01)
Denominator used	Expected stays (calculated by age-sex standardisation)
Data source of denominator	ISD SMR01, and NRS mid-year resident population estimates.
Method of construction of indicator	For each data zone, the sum of observed CISs by sex and 5-year age band was divided by the corresponding sum of expected CISs. For further details, please see methodology section (2.3) of SIMD 2006 technical report[8].
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.

[8] http://www.scotland.gov.uk./Publications/2006/10/13142913/0

Implications of comparing this indicator with the one used in SIMD 2016	Indirectly standardised ratios have limitations for drawing comparisons. In comparing one data zone between different SIMD releases, if it had a ratio of 100 in SIMD16 and 110 in SIMD 2020, you could say that the data zone has got worse **relative to Scotland**. However, you cannot say whether the data zone has actually worsened – only that it has worsened relative to Scotland. In comparing two data zones for the same SIMD release, if one data zone had a ratio of 110 and another a ratio of 115, you could say that both data zones are worse than the Scottish average (100), but for statistical reasons you **cannot** say that the data zone with the ratio of 115 is worse than the one with the ratio of 110.
Other data quality issues	In a number of areas close to the English border, Scottish residents may be using health services provided by NHS England. These cases are not captured in the Scottish data. For SIMD16, ratios for three data zones in Dumfries and Galloway were statistically adjusted to account for these cases. After consultation with ISD and NHS Dumfries and Galloway, the decision was taken not to adjust figures for SIMD 2020 as it was not possible to obtain detailed information on the number of affected cases in NHS Dumfries and Galloway and NHS Borders, and to ensure consistency with data published by ISD, which do not include cases of cross-border access to health services.
Disclosure control	N/A
Geo-referencing	All postcodes are validated at source. Data exclude people where no match to a data zone was possible (e.g. homeless, incomplete postcode information).
Availability of data	The standardised ratio is available on www.gov.scot/SIMD.

Proportion of population being prescribed drugs for anxiety, depression or psychosis

General description of indicator	Proportion of patients being prescribed anxiolytic, antipsychotic or antidepressant drugs. Derived from paid prescriptions data at patient level.
Indicator type	Proportion
Time period	Financial year 2017-18
Data source	ISD Prescribing Information System (PIS) All data held in PIS is sourced from Practitioner Services Division (PSD) within NHS National Services Scotland, who are responsible for the remuneration and reimbursement of dispensing contractors within Scotland.
Denominator used	NRS mid-2017 population estimates
Data source of denominator	NRS
Method of construction of indicator	The prescribing team within ISD maintains a detailed database of all NHS prescriptions dispensed in the community in Scotland. Anxiolytic, antipsychotic and antidepressant drugs can be identified through the British National Formulary (BNF) Codes 4.1.2 (anxiolytics), 4.2 (antipsychotics), and 4.3 (antidepressants). **Please note** that previous releases of SIMD have used existing BNF section classifications for reporting of drugs used in mental health. Since this indicator was last published for SIMD16, the structure of BNF medicines classification has changed and the section descriptions used are no longer applicable to the new structure. In order to maintain consistency and comparability with previous years, this report uses the “legacy BNF” structure, consistent with what has been used in previous years. Up to date information on the availability and therapeutic uses of medicines can be found on the BNF website. Patient-based analysis is possible because comprehensive patient identifiable data are available in the prescribing dataset. All NHS patients have a unique Community Health Index (CHI) number; this makes it possible to identify which prescription items have been dispensed for individual patients. Prior to April 2009, the proportion of prescriptions with a valid CHI number recorded was not high enough to make patient-based analysis possible. For medicines used in mental health

	the CHI capture/completeness rates are now high enough to permit accurate patient analyses. In 2017-18, the CHI capture rate for anxiolytic, antipsychotic and antidepressant drugs was 98.4%. The data zone attributable to each patient with a valid CHI number has been derived from the address recorded on CHI, as at the time that the latest prescription for anxiolytic, antipsychotic and antidepressant drugs in 2017-18 was processed by the Data Capture Validation and Pricing system in PSD. The proportion of the population being prescribed drugs for anxiety, depression or psychosis in each data zone was estimated by dividing the aggregated count of patients per data zone by the data zone population estimates for mid-2017.
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	This indicator remains the same as in SIMD16.
Other data quality issues	There is an issue with extracting data from PIS by patient geography which NSS IT continues to work on resolving. An estimate of up to 1-2% of records may have an out of date postcode. This issue does not appear to affect any one geographic area more than any other. The CHI capture rate for these medications is 98%. Figures are not adjusted to account for this.
Disclosure control	N/A
Geo-referencing	All postcodes are validated at source. Data excludes people where no match to a data zone was possible (e.g. homeless, incomplete or redundant postcode information).
Availability of data	The proportion of the population being prescribed drugs for anxiety, depression or psychosis is available on www.gov.scot/SIMD.

Proportion of live singleton births of low birth weight

General description of indicator	Proportion of live singleton births with low birth weight (less than 2,500 grams), where birth figures exclude home births and births at non-NHS hospitals. A singleton is a baby from a pregnancy resulting in only one live or stillbirth. No account is taken of gestational age. The year shown refers to the year of discharge from hospital.
Indicator type	Proportion
Time period	Financial years 2014-15 to 2017-18 (four-year aggregate)
Data source	NHS Scotland Information Services Division (ISD), Scottish Morbidity Record (maternity records) (SMR02)
Denominator used	All live singleton births with a known birth weight
Data source of denominator	ISD SMR02 (note that this excludes home births and births at non-NHS hospitals)
Method of construction of indicator	For each data zone, the number of low birth weight live singleton births was divided by the total number of live singleton births.
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16. However, as these are very small areas, be aware of the effect of small numbers and potential zeros.
Implications of comparing this indicator with the one used in SIMD 2016	See comment above. Similarly, please take care in drawing conclusions about time trends, as the proportions by data zone are generally small and not robust.
Other data quality issues	N/A
Disclosure control	N/A
Geo-referencing	All postcodes are validated at source. Data excludes births where no match to a data zone was possible (e.g. homeless, incomplete postcode information).
Availability of data	The proportion of live singleton births of low birth weight is available on www.gov.scot/SIMD.

Education, skills and training domain

SIMD 2020 indicators	2020 weight	SIMD16 indicators	2016 weight	Summary of change
School pupil attendance Source: SG, 2014-15 and 2016-17	0.27	School pupil attendance SG, 2012-13 and 2014-15	0.21	No change
Attainment of school leavers SG, 2015-16 to 2017-18	0.20	Attainment of school leavers SG, 2011-12 to 2014-15	0.23	No change
Working age people with no qualifications: standardised ratio 2011 Census	0.29	Working age people with no qualifications: standardised ratio 2011 Census	0.29	No change
17-21 year olds enrolling into higher education HESA, 2015-16 to 2017-18	0.14	17-21 year olds enrolling into higher education HESA 2012-13 to 2014-15	0.15	No change
People aged 16-19 not participating in education, employment or training SDS, 2015-16 to 2017-18	0.10	People aged 16-19 not in full time education, employment or training (NEET) School Leavers 2012-13 to 2013-14, DWP 2013 and 2014	0.12	Skills Development Scotland annual participation measure replaces DWP NEET measure.

Notes on the construction of the education domain

Missing data

For some indicators, the population of the considered age group was zero in some data zones during the time period considered, and indicator rates and ranks could not be determined. Missing rates and ranks are denoted by '*'.

To calculate the overall rankings for the education domain, the normalised scores for these data zones were set to zero before combining the indicators. As a result, the indicators with missing values moved the overall domain ranking of these data zones towards a middle ranking.

Weights

The weightings of the individual indicators in the education domain are rounded and may therefore not add up to 100 per cent.

Age-sex standardisation

The aim of standardisation is to provide a summary ‘adjusted’ rate to take into account underlying differences (such as age and sex) of a study population relative to a ‘reference’ population.

Indirect age-sex standardisation is based on a comparison of observed to expected numbers of cases by applying age-specific rates from a ‘standard population’ to the population of interest. For example, if the study population is the resident population of a data zone, then the standard population used here is the Scotland total.

For more information on direct and indirect standardisation techniques, please refer to the NHS guidance[9].

The following tables give more detailed information on each of the indicators in the education, skills and training domain used in constructing SIMD 2020.

[9] https://www.scotpho.org.uk/media/1400/phi-standardisation-guidance-v21.docx

School pupil attendance

General description of indicator	Percentage of pupils who attend school 90% or more of the time for each data zone in Scotland; includes pupils who attend publicly funded primary, secondary and special schools.
Indicator type	Percentage of pupils with attendance of 90% or above over a two year period
Time period	School years 2014-15 and 2016-17 (no data collection in 2015-16).
Data source	Data are provided to the Scottish Government by the local authorities and managers of mainstream grant-aided schools
Denominator used	Total number of pupils
Data source of denominator	Data are provided to the Scottish Government by the local authorities and managers of mainstream grant-aided schools.
Method of construction of indicator	Information is received at pupil-level. Data are aggregated over two years, with the intention of reducing the impact of fluctuations from one year to the next due to a small number of pupils per data zone.
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	This indicator can be sensitive to small numbers in some data zones where the number of school-age children is small.
Other data quality issues	N/A
Disclosure control	Values have been supressed where there are fewer than nine pupil records in a data zone that are either above or below the 90% attendance threshold.
Geo-referencing	Pupil home postcode is one of the indicators collected for each pupil in the annual pupil census. Postcodes were known for over 99% of pupils in publicly funded schools. Data used in the formulation of this indicator do not include pupils with a missing or invalid postcode.

Availability of data	Percentages are available at www.gov.scot/SIMD. National data are published in the Scottish Government statistical publication[10].

[10] https://www2.gov.scot/Topics/Statistics/Browse/School-Education/Summarystatsforschools

Attainment of school leavers

General description of indicator	Score based on school leavers' highest level of the qualification for pupils who attend publicly funded secondary schools.
Indicator type	Average score (three year average)
Time period	School years 2015-16 to 2017-18
Data source	Exam results data on national qualifications are obtained from the Scottish Qualifications Authority (SQA) and are post-review data. These data are linked to the pupil information from school leavers dataset, using the pupil's Scottish candidate number from the pupil census. For more information please see the background notes to the publication 'School leaver attainment and initial destinations', available here: https://www.gov.scot/publications/summary-statistics-attainment-initial-leaver-destinations-1-2019-edition/
Denominator used	Total number of school leavers
Data source of denominator	School leavers dataset
Method of construction of indicator	Using the Scottish Credits and Qualifications Framework for school leavers, the score is calculated by identifying the best level of qualification each pupil leaves school with. The score is calculated by multiplying the highest qualification level achieved by each pupil by a corresponding factor. Level 3 qualifications are multiplied by three, level 4 by four, level 5 by five, level 6 by six and level 7 by seven. This indicator looks at the highest qualification attained by each pupil, not the number of qualifications attained. For example, one pupil who leaves school with four level 3 qualifications will score three, whilst a pupil leaving school with one level 5 qualification will score five. The total score is then divided by the total number of school leavers in each geographical area. Data are based on an average of three years and include all school leavers in secondary schools and special schools. Data from independent schools are not included. This is because there is a lack of data zone level information and independent schools may choose to sit GCSEs or other English qualifications for which we have no information. We do not hold Scottish candidate numbers of independent school pupils. The number of pupils on the roll is taken from the September pupil census for the relevant year, which was carried out

	through ScotXed project. The year shown relates to the academic year: http://www.gov.scot/Topics/Statistics/ScotXed
Key decisions on methodology	Indicator methodology remains the same as SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	This indicator remains the same as in SIMD16.
Other data quality issues	N/A
Disclosure control	In some data zones there were fewer than five school leavers in the school years considered, therefore the rate has been suppressed for publication. This is denoted by "*". Unsuppressed figures were used in the SIMD calculations.
Geo-referencing	A pupil is assigned to a geographical area based on their home address. If their postcode is missing their school's postcode is used. Valid postcodes were submitted for about 99% of pupils in publicly funded secondary schools.
Availability of data	This indicator is available for download at www.gov.scot/SIMD.

Working age adults with no qualifications: standardised ratio

General description of indicator	The indicator shows the percentage of working age adults (aged 25-64) that were recorded as having no qualifications in the 2011 Census.
Indicator type	Indirectly standardised ratio.
Time period	2011
Data source	NRS, 2011 Census
Denominator used	Expected number of people with no qualifications (calculated using indirect age-sex standardisation)
Data source of denominator	2011 Census
Method of construction of indicator	The observed number of people with no qualifications was divided by the expected number for each data zone to provide a ratio.
Key decisions on methodology	The official source for data on the proportion of people with no qualifications is the Labour Force Survey (LFS). As the LFS is a survey, reliable data are not available at data zone level. For this reason the only viable source for this indicator at data zone level is the 2011 Census. Standardised ratios are used in order to account for known cohort effects in the likelihood of leaving school with no qualifications.
Comparison with 2016 indicator	Indicator methodology is the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	This indicator is the same as in SIMD16.
Other data quality issues	N/A

Disclosure control	This indicator may have been disclosure controlled using either or both of these techniques: - targeted record swapping; - restriction of detail, particularly at low level geographies Visit the 'National Archives' website for further details[11].
Geo-referencing	No geo-referencing issues
Availability of data	The standardised ratio can be downloaded at www.gov.scot/SIMD. Census data are also published by NRS[12].

[11] http://webarchive.nationalarchives.gov.uk/20160105160709/http:// ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/methods/statistical-disclosure-control-methods/index.html

[12] http://www.scotlandscensus.gov.uk/

People aged 17-21 year olds entering university

General description of indicator	The indicator reflects the enrolment rate based on the number of 17-21 entrants to first degree courses domiciled in each data zone before the start of their course, compared to the total number of 17-21 year olds resident in that data zone over the same period.
Indicator type	Rate (over three years).
Time period	2015-16 to 2017-18
Data source	Higher Education Statistics Agency
Denominator used	NRS mid-2015 to 2017 population estimates of 17-21 year olds, adjusted to account for large student populations (see note on data quality below).
Data source of denominator	NRS
Method of construction of indicator	The number of 17-21 year olds who entered a first degree programme at a UK university between 2015-16 and 2017-18 from each data zone was divided by the total population estimate of 17-21 year olds resident in that data zone in the same time period. This gives a rate of 17-21 year olds in each data zone enrolling for a first degree course at a UK university.
Key decisions on methodology	Indicator methodology remains the same as SIMD16. Entrants to first degree courses at UK universities using data supplied by HESA was used as the numerator for SIMD 2020. Study at degree level has been chosen, as this level provides the highest gains in future earning potential and reduces double counting of students who progress from further education to higher education (such as from an HND to a university degree course).
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	This indicator remains the same as in SIMD16.

Other data quality issues	The count of students enrolling relates to home address before study. An adjustment is therefore required for the denominator population as this includes students at their term-time address which affects a small number of data zones containing student halls. The quality assurance process identified that small absolute changes to the average 16 year old populations between the time periods for SIMD 2016 and SIMD 2020 resulted in substantial changes to adjusted denominators and rates. The adjustment methodology was reviewed, concluding that the use of the average 16 year old population concurrent with the higher education denominator time period was not representative of the true 17-21 year old population. Small fluctuations in population estimates are likely to introduce volatility and inaccuracy to the indicator. Based on these findings, a new method was developed which uses the weighted averages of all 16 year old population estimates that are relevant to the higher education denominator. Individuals who were in the SIMD 2020 denominator would have been 16 years old between 2010 and 2016 with cohorts of individuals being included in the denominator on consecutive years. The new method uses population estimate averages weighted according to how many times they would be included in the denominator. For example, those who were 16 years old in 2014 would have been 17 years old in 2015, 18 years old in 2016, and 19 years old in 2017: the 2014 population estimate is therefore given a weight of three. It is noted that some 16 year olds will have moved to another area before the relevant time period, however the numbers involved and their effect on the indicator are assumed to be small. In some data zones, there were no 17-21 year olds in the years considered, therefore a rate could not be determined. This is denoted by '*'.
Disclosure control	N/A
Geo-referencing	Data exclude people where no match to a data zone was possible (e.g. homeless, incomplete postcode information). This affects 0.6% of the students.
Availability of data	This indicator can be downloaded at www.gov.scot/SIMD.

People aged 16-19 not participating in education, employment or training

General description of indicator	The indicator shows the proportion of 16 to 19 year-olds who are not participating in education, employment or training
Indicator type	Percentage (over three years)
Time period	Financial years 2015-16 to 2017-18
Data source	Data provided by Skills Development Scotland from the annual participation measure dataset[13]. The APM dataset uses data from the 'Opportunities for All' dataset managed by SDS. This dataset includes individual-level information from key partners, including local authorities, colleges, the Department for Work and Pensions and Student Awards Agency for Scotland.
Denominator used	APM cohort size over three years (2016-2018)
Data source of denominator	SDS
Method of construction of indicator	Each individual in the APM dataset is categorised as participating or not participating in education, training or employment, or has an unconfirmed status. The status of each individual is identified by the category that accounts for the greatest number of days over the course of each year. For each data zone the numbers of individuals categorised as not participating in education, employment or training in each of the three years of data were combined and that figure was divided by the combined APM cohort size across the three years.
Key decisions on methodology	Data were aggregated over three years, with the intention of reducing the impact of fluctuations from one year to the next due to small numbers.
Comparison with 2016 indicator	This indicator replaces the SIMD16 not in education, employment or training indicator.
Implications of comparing this indicator with the one used in SIMD 2016	This is a new indicator and cannot be directly compared with the SIMD16 indicator.

[13] https://www.skillsdevelopmentscotland.co.uk/publications-statistics/statistics/annual-participation-measure/?page=1&statisticCategoryId=7&order=date-desc

Other data quality issues	SDS receives data from DWP for those aged 18 and over in receipt of Job Seekers Allowance, Employment and Support Allowance or Income Support. These benefits have been gradually replaced by Universal Credit (UC) since March 2016. At present, SDS does not receive data from DWP on UC claimants, and therefore is not able to systematically identify or report on 18 or 19 year olds who are unemployed and claiming out of work benefits.
Disclosure control	N/A
Geo-referencing	Data zones are derived from an individual's home postcode at the end of each financial year. All individuals with a valid postcode were included for this indicator. This approach differs from that used in SDS publications to ensure consistency for sub-local authority reporting, which does not include individuals whose reporting local authority is different from their home postcode local authority. For example, an individual could attend or have attended a school in one local authority while living in a neighbouring authority.
Availability of data	The percentage can be downloaded at www.gov.scot/SIMD.

Geographic access to services domain

SIMD 2020 indicators	2020 weight	SIMD16 indicators	2016 weight	Summary of change
Drive time sub-domain (weight in access domain: 60%)				
to GP surgery	0.23	to GP surgery	0.24	No changes
to post office	0.14	to post office	0.15	No changes
to retail centre	0.24	to retail centre	0.21	No changes
to primary school	0.11	to primary school	0.11	No changes
to secondary school	0.16	to secondary school	0.15	No changes
to petrol station	0.13	to petrol station	0.14	No changes
Public transport sub-domain (weight in access domain: 30%)				
to GP surgery	0.45	to GP surgery	0.47	No changes
to post office	0.28	to post office	0.28	No changes
to retail centre	0.27	to retail centre	0.25	No changes
Broadband sub-domain (weight in access domain: 10%)				
Percentage of premises that do not have access to superfast broadband.				New indicator for SIMD 2020

Notes on the construction of the access domain

Weighting

The drive time sub-domain scores receive a higher weight in the overall access domain because the data is more robust and consistent across Scotland. The domain takes into account access to six services whereas the public transport sub-domain only takes into account three different services. As with SIMD16, it was not possible to include public transport times to primary and secondary schools as there is no consistent time table information available on school buses.

Changes to the data

The data source for the primary and secondary school locations has changed from Ordnance Survey PointX to the Scottish Government School Roll. The speed data used in SIMD 2020 is based on TRAFFICMASTER rather than the INRIX product used for SIMD16. The decision to use TRAFFICMASTER for SIMD 2020 was based on advice from Basemap, who noted that TRAFFICMASTER is better suited to the OS Highways product within TRACC (the software used to calculate journey times).

Detailed information on the background and methodology for the access domain can be found in Annex B.

Drive time sub-domain

- Average drive time to a primary school
- Average drive time to a secondary school
- Average drive time to a GP surgery
- Average drive time to a post office
- Average drive time to a retail centre
- Average drive time to a petrol station

General description of indicator	Population weighted average time taken to reach key services by driving
Indicator type	Population weighted average drive time in minutes
Time period	2018
Data sources	Road network: Ordnance Survey Highways, October 2018 Road speed: TRAFFICMASTER speed data, October 2018 Public Transport timetable / route data: Traveline National Dataset and Associate of Train Operating Companies data (October 2018) Service locations: Primary schools and secondary schools: Scottish Government School Roll, October 2018 Petrol stations and GP surgeries: Ordnance Survey PointX, September 2018 Retail centres: CACI Retail, 2018 Post Offices : Royal Mail, 2018 plus Ordnance Survey PointX, September 2018 Origin locations : Census Output Area population-weighted centroids: National Records of Scotland, 2011 Census
Denominators used	2016-18 primary school population (for drive time to primary schools) 2016-18 secondary school population (for drive time to secondary schools) 2011 Census total population (for all other indicators)
Data sources of denominators	Scottish Government pupil census, 2016-18 National Records of Scotland, 2011 Census

Method of construction of indicator	For each census output area, the closest trip to each destination was calculated. The results were aggregated to data zone level to provide an average drive time for each data zone. Further details are available in the GI-SAT report on the construction of the SIMD 2020 geographic access domain in Annex B.
Key decisions on methodology	Drive times were calculated using Basemap TRACC software. The software searches for the closest of each type of destination to every origin, and calculates its drive time. Where drive times may include a ferry journey, the calculation was schedule-based (same day/times as public transport times – see below) and averaged in order to take account of the actual ferry timetable data and variability of services throughout the day. Thirty minute wait times at ferry terminals were included. Passenger ferry routes were excluded from drive time calculations. The origin of any trip is the population weighted census output area centroid. Road speeds are based on observed road speeds at different times of the day. Post Office locations reflect post office services at a fixed location available for 6 hours or more per week. Primary and secondary school populations were used to weight primary and secondary school drive times.
Comparison with 2016 indicator	The services used are the same as for SIMD 2016, however there was a change to the source data for the primary and secondary school destinations.
Implications of comparing this indicator with the one used in SIMD 2016	The change to the source data for the primary and secondary school locations has had a very small impact on the results; rather than the points being 'snapped' to the nearest road segment, they are located on the buildings, meaning the software includes a very short walk (around 20 to 30 seconds) to get from the road to the building as part of the journey time. In SIMD 2016 the school journeys ended on the road segments.

Other data quality issues	Drive times have been imputed to 190 minutes for at least some output areas in each of the following eleven data zones (for one or more services) as there was either no car ferry connection within the time window considered, or the journey was longer than the maximum journey time of 180 minutes: S01007284 Mull, Iona, Coll and Tiree (retail centre, secondary school) S01007287 Mull, Iona, Coll and Tiree (GP, retail centre, petrol station, secondary school) S01007289 Oban South (all services) S01007310 Loch Awe – 03 (all services) S01007324 Whisky Isles (retail centre, secondary school) S01010504 Lochaber West (GP, retail centre, petrol station, secondary school) S01010506 Lochaber West (GP, retail centre, petrol station, secondary school) S01011831 Isles (Orkney) (retail centre, petrol station, primary school, secondary school) S01011832 Isles (Orkney) (retail centre, secondary school) S01012387 Shetland South (GP, retail centre, petrol station, secondary school) S01012416 North and East Isles (Shetland) (GP, retail centre, petrol station, primary school, secondary school)
Disclosure control	N/A
Geo-referencing	All service locations were geo-referenced.
Availability of data	Drive times can be downloaded on a data zone level at www.gov.scot/SIMD.

Public transport sub-domain

- Public transport travel time to a retail centre
- Public transport travel time to a GP surgery
- Public transport travel time to a post office

General description of indicator	Population weighted average travel times to selected services using public transport (bus, rail, coach, ferry and tram) and/or walking.
Indicator type	Population weighted average public transport travel time in minutes
Time period	2018
Data sources	Road network: Ordnance Survey Highways, October 2018 Public Transport timetable / route data: Traveline National Dataset and Associate of Train Operating Companies data (October 2018) Service locations: GP surgeries: Ordnance Survey PointX, September 2018 Retail centres: CACI Retail, 2018 Post Offices : Royal Mail, 2018 plus Ordnance Survey PointX, September 2018 Origin locations : Census Output Area population-weighted centroids: National Records of Scotland, 2011 Census
Denominator used	2011 Census total population
Data source of denominator	National Records of Scotland, 2011 Census
Method of construction of indicator	For each census output area, the closest trip to each destination was calculated. The results were aggregated to data zone level to provide an average public transport time for each data zone. Further details are available in the GI-SAT report on the construction of the SIMD 2020 geographic access domain in Annex B.

Key decisions on methodology	Average public transport travel time calculated at Census output area level from four trips (two outbound and two inbound) throughout the day in order to reflect changes in public transport availability throughout the day. Outbound times calculated for Tuesday 7 am – 11 am and 12 pm – 4 pm. Inbound times calculated for Tuesday 12 pm – 4 pm and 5 pm – 9 pm. A three hour maximum journey time cut off was applied to all trips. Twenty minute wait times at ferry terminals were included where journeys included a ferry trip. Post Office locations reflect post office services at a fixed location available for six hours or more per week.
Comparison with 2016 indicator	The services used are the same as for SIMD 2016, however there was a change to the source data for the primary and secondary school destinations.
Implications of comparing this indicator with the one used in SIMD 2016	The change to the source data for the primary and secondary school locations has had a very small impact on the results; rather than the points being 'snapped' to the nearest road segment, they are located on the buildings, meaning the software includes a very short walk (around 20 to 30 seconds) to get from the road to the building as part of the journey time. In SIMD 2016 the school journeys ended on the road segments.
Other data quality issues	For some data zones, where a public transport time could not be calculated for output areas within the data zone, travel times were estimated based on the drive times.
Disclosure control	N/A
Geo-referencing	All service locations were geo-referenced.
Availability of data	Public transport travel times can be downloaded on a data zone level at www.gov.scot/SIMD.

Broadband sub-domain

General description of indicator	The indicator shows the percentage of premises that do not have access to superfast broadband, defined as providing a minimum 30Mbit/s download speed. This includes properties with no broadband coverage.
Indicator type	Percentage
Time period	January 2019
Data sources	The Ofcom Connected Nations update: Spring 2019 provides coverage and service availability information received from communications providers as of January 2019[14].
Denominator used	Total premises.
Data source of denominator	Ofcom Connected Nations update: Spring 2019.
Method of construction of indicator	The *Connected Nations 2018 methodology report*[15] explains how broadband operators provide data for each address where a service is provided. Note that the Ofcom *Connected Nations* report covers 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 such as city centres.
Key decisions on methodology	N/A
Comparison with 2016 indicator	This indicator is new for SIMD 2020
Implications of comparing this indicator with the one used in SIMD 2016	New indicator.

[14] https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-update-spring-2019.

[15] https://www.ofcom.org.uk/data/assets/pdffile/0033/129975/connected-nations-2018-methodology.pdf.

Other data quality issues	Due to variations in broadband performance over time, these data should not be regarded as a definitive view of fixed broadband infrastructure. However, the information is useful in identifying variations in broadband availability.
Disclosure control	N/A
Geo-referencing	Data are provided for 2011 Census output area and aggregated to data zone level.
Availability of data	Percentages of premises with no access to superfast broadband can be downloaded at data zone level at www.gov.scot/SIMD.

Crime domain

SIMD 2020 indicators	SIMD16 indicators	Summary of change
Recorded crimes of violence Source for all: Police Scotland 2017-18	Recorded crimes of violence Source for all: Police Forces, 2014-15	No change
Recorded sexual offences	Recorded sexual offences	No change
Recorded domestic housebreaking	Recorded domestic housebreaking	No change
Recorded vandalism	Recorded vandalism	No change
Recorded drugs offences	Recorded drugs offences	No change
Recorded common assault	Recorded common assault	No change

Notes on the construction of the crime domain

Consistency with published sources

The current methodology consists of calculating the recorded crime rate for selected crime types per 10,000 population. Recorded crime counts are based on geo-referenced data for 2017-18 provided by Police Scotland.

Police stations

In order to reduce bias against areas that contain a police station, crimes that have been identified as being recorded within 50 metres of the centre of a police station have been excluded.

Weighting of crime types

There is no standard methodology to differentiate between the severity of different crime types, therefore the rate has been calculated from the unweighted total counts.

Crimes excluded from the crime domain

The SIMD crime domain does not include all types of crime or offence recorded by the police. Certain crimes have been excluded because of data quality issues, or because they are less meaningful in terms of deprivation at a neighbourhood level.

For example, some crime types were excluded because they are directed at businesses and/or concentrated in centres of retail activity rather than neighbourhoods – examples include shoplifting and non-domestic housebreaking. Other crime types are harder to locate geographically – examples include fraud and speeding offences.

Other crimes might be thought suitable for exclusion, but where such crimes formed part of a recognised group (e.g. violence, vandalism) and accounted for very small numbers, they were not explicitly excluded.

Geographical coverage

Police Scotland provided all available geographical/address information for each crime recorded, including postcodes, map references and address details. The exact postcode is not necessarily required to geo-reference a crime as long as there is sufficient information to allocate the crime to a data zone. Where a force could only provide address details, the Scottish Government imputed postcodes and/or map references so that data zones could be identified for the majority of the crimes.

Crimes that could not be geo-referenced were allocated to the council area based on the information provided. The remaining SIMD crime count was grossed up to the council area total.

Population denominators

The population denominators used are the National Records for Scotland mid-2017 small area estimates.

There may be some bias for data zones that fall in town centres, where the resident population may be small compared with the daytime or night-time populations. Day and night populations are not available at the small area level, and so total resident population has been used.

Special events, such as music festivals or sporting events, may also result in temporary increases in population accompanied by an increase in the number of recorded crimes, which are not reflected in the small area resident population estimates. Such temporary changes to the population of a data zone should be taken into account when comparing SIMD crime rates in different areas.

Missing data

In a small number of data zones, the population was zero in the year considered, and crime rates could not be determined. Missing rates are denoted by '*'.

Comparison with National Statistics on recorded crime

The number of crimes included in the SIMD crime domain is taken from a live operational database which is maintained by Police Scotland. Should the initial report of an incident suggest some form of criminality may have taken place, a crime record will be created on Police Scotland's database. These crime records can then be amended as investigations continue into the incident. These amendments tend to reduce the total volume of crime within any time period as the most common change is the 'declassification' of some crime records (known as 'no-criming') after further investigations conclude no criminality has occurred.

An example of this could be where a report alleging vandalism of property turns out to have been accidental damage. As a result of declassifications, and other administrative changes to crime records, the number of crimes for any given period would differ depending on when the information was extracted. Therefore we would expect there to be a slight difference between the SIMD crime domain and the figures presented in the *Recorded crime in Scotland* National Statistics bulletin for the same year.

The following tables give more detailed information on the crime indicators and the constituent offences which are used in constructing SIMD 2020.

Recorded crimes

General description of indicator	Recorded crime rate of selected crimes of violence, sexual offences, domestic housebreaking, vandalism, drug offences and common assault. The overall indicator is a sum of each SIMD crime per 10,000 population.
Indicator type	Rate per 10,000 people
Time period	2017-18
Data source	Police Scotland
Denominator used	Total Population 2017
Data source of denominator	NRS 2017 mid-year population estimates
Method of construction of indicator	Crimes happening within 50 metres of a police station were removed from the data. This was to ensure that crimes recorded as happening in a police station did not mask the level of crime happening in the neighbourhood of the police station. The total SIMD crime rate was constructed by summing the selected crime counts and dividing the total by the total data zone population and multiplying by 10,000 to obtain the rate per 10,000 population.
Key decisions on methodology	The indicators remain the same as those used in SIMD16.
Comparison with 2016 indicator	This indicator remains the same as in SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	This indicator remains the same as in SIMD16.
Other data quality issues	When looking at crime domain ranks and rates of SIMD crimes, the type of area and resident population should be considered. Some areas, such as town centres and areas around football stadia, will see large numbers of people in an area at a particular time of day or day of the week or year and an associate increase in crime. Crime rates were rounded to the nearest integer after grossing the counts (see geo-referencing).
Disclosure control	Counts for data zones where the crime count is less than three have been suppressed.

Geo-referencing	There was a lack of full address information for some crimes/ offences. In these cases, partial addresses were geo-referenced manually. Crimes that could not be geo-referenced were allocated to the council area based on the information provided. The remaining SIMD crime count was grossed up to the council area total based on the information provided.
Availability of data	The crime counts and rates can be downloaded at www.gov.scot/SIMD.

SIMD crime codes

The crime domain consists of categories of recorded crimes or offences. Certain crimes which form part of recognised groups, such as violence, and account for very small numbers, are included for ease of comparison with published totals, even though they may not seem entirely relevant to SIMD.

Crime group name	Crime code	Crime / offence name
Crimes of violence	1/000	Murder
	2/000	Attempted murder
	3/001	Culpable homicide (common law)
	3/002	Causing death by dangerous driving
	3/003	Death by careless driving under the influence of drink or drugs
	3/004	Causing death by careless driving
	3/005	Illegal driver, disqualified/unlicensed etc involved in fatal accident
	3/006	Corporate homicide
	4/000	Serious assault
	4/001	Causing serious injury etc. by culpable and reckless conduct
	6/000	Robbery and assault with intent to rob
	7/000	Threats and extortion
	8/001	Cruelty to and unnatural treatment of children
	8/002	Child stealing (plagium)
	8/003	Exposing child under 7 to risk of burning
	9/000	Abortion
	10/000	Concealment of pregnancy
	11/001	Possess a firearm with intent to endanger life, commit crime etc.
	11/002	Abduction
	11/003	Ill treatment of mental patients
	11/004	Cruel and unnatural treatment of an adult
	11/005	Drugging
	11/006	Chemical weapon offences
	11/007	Female genital mutilation
Sexual offences	13/001	Illegal homosexual acts
	13/002 and 18/021	Bestiality (change to crime code only)
	13/003	Assault to commit unnatural crimes
	14/000	Rape (offence prior to 1 December 2010)
	14/001	Rape of male (16+)
	14/002	Rape of female (16+)
	14/003	Rape of older male child (13-15 years)
	14/004	Rape of older female child (13-15 years)
	14/005	Rape of young male child (Under 13)
	14/006	Rape of young female child (Under 13)

	15/000	Attempted rape (Offence prior to 1 December 2010)
	15/001	Attempted rape male (16+)
	15/002	Attempted rape female (16+)
	15/003	Attempted rape older male child (13-15)
	15/004	Attempted rape older female child (13-15)
	15/005	Attempted rape young male child (under 13)
	15/006	Attempted rape young female child (under 13)
Crimes of dishonesty	19/004	Theft by housebreaking domestic property (dwelling)
	19/005	Theft by housebreaking domestic property (non-dwelling)
	19/007	Housebreaking with intent to steal domestic property (dwelling)
	19/008	Housebreaking with intent to steal domestic property (non-dwelling)
	19/010	Attempted housebreaking with intent to steal domestic property (dwelling)
	19/011	Attempted housebreaking with intent to steal domestic property (non-dwelling)
Vandalism	32/001	Fire-raising excluding muirburn
	32/003	Muirburn
	33/001	Vandalism, reckless damage and malicious mischief
	33/002	Reckless conduct with firearms
	33/003	Flying aircraft to the danger of life or property
	33/004	Endangering rail passengers
	33/005	Reckless driving at common law
	33/006	Culpable neglect of duty
	33/007	Endangering ship by breach of duty, obtain ship by misrepresentation
	33/010	Computer Misuse Act 1990
	33/011	Culpable and reckless conduct (not with firearms)
	33/012	Vandalism
	33/013	Reckless damage
	33/014	Malicious mischief
Drug offences	44/001	Illegal importation of drugs
	44/002	Production, manufacture or cultivation of drugs
	44/003	Supply, possession with intent to supply etc of drugs
	44/004	Possession of drugs
	44/005	Drugs, money-laundering offences
	44/099	Drugs, other offences
Common assault	47/001	Common assault
	47/006	Common assault of an emergency worker

Housing domain

SIMD 2020 indicators	SIMD16 indicators	Summary of change
Persons in households that are overcrowded Source: 2011 Census	Persons in households that are overcrowded Source: 2011 Census	No change
Persons in households without central heating Source: 2011 Census	Persons in households without central heating Source: 2011 Census	No change

Notes on the construction of the housing domain

The SIMD 2020 housing domain uses 2011 Census data, and is constructed by summing the two indicators. Although there is a chance of some overlap between indicators, this has been accepted as reflecting that people living in a household with both of these attributes are more deprived than those with only one.

Persons in households that are overcrowded

General description of indicator	This indicator provides a measure of material living standards and gives the proportion of household population that live in overcrowded housing based on the occupancy rating. This compares the actual number of rooms in the house to the number of rooms which are required by the household, based on the relationships between them and their ages. Overcrowding is defined to mean households with an occupancy rating of -1 or less. This means that there is at least one room too few in the household.
Indicator type	Percentage of household population
Time period	2011
Data source	NRS, 2011 Census
Denominator used	2011 Census household population
Data source of denominator	NRS, 2011 Census
Method of construction of indicator	The indicator is a simple proportion of the number of people living in overcrowded households, divided by the 2011 household population.
Key decisions on methodology	The methodology used to construct the indicator is the same as that used for previous editions of SIMD.
Comparison with 2016 indicator	Same indicator as for SIMD16.
Implications of comparing this indicator with the one used in SIMD 2016	
Other data quality issues	There has been some criticism that the Census measure of overcrowding (the occupancy rating) is not as sophisticated as the 'bedroom standard' which is generally used to assess overcrowding, and tends to overstate. For example, on the Census definition a one person household can have an occupancy rating of -1.
Disclosure control	This indicator may have been disclosure controlled using either or both of these techniques: - targeted record swapping; - restriction of detail particularly at low level geographies

	Visit the 'National Archives' website for further details.[16]
Geo-referencing	No geo-referencing issues
Availability of data	The data is available on the SIMD website www.gov.scot/SIMD. 2011 Census data is available on the Scottish Census website[17].

[16] http://webarchive.nationalarchives.gov.uk/20160105160709/http:// ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/methods/statistical-disclosure-control-methods/index.html

[17] http://www.scotlandscensus.gov.uk/

Persons in households without central heating

General description of indicator	This indicator provides a measure of material living standards and gives the proportion of household population that live in a house that is centrally heated. A household's accommodation is described as 'with central heating' if it has central heating in some or all rooms (whether used or not). Central heating includes gas, oil or solid fuel central heating, storage heaters, and solar heating.
Indicator type	Percentage of household population
Time period	2011
Data source	NRS, 2011 Census
Denominator used	2011 Census household population
Data source of denominator	NRS, 2011 Census
Method of construction of indicator	The indicator is a simple proportion of the number of people living in households with no central heating, divided by the 2011 household population.
Key decisions on methodology	The methodology used to construct the indicator was the same as that used for previous editions of SIMD.
Comparison with 2016 indicator	This is the same indicator as for SIMD16.
Implications of comparing this indicator with the one used in SIMD 2012	Data zone level indicator scores cannot be compared due to changed data zone boundaries.
Other data quality issues	The Census question on central heating does not distinguish between whole and partial house central heating.
Disclosure control	This indicator may have been disclosure controlled using either or both of these techniques: - targeted record swapping; - restriction of detail particularly at low level geographies Visit the 'National Archives' website for further details.[18]

[18] http://webarchive.nationalarchives.gov.uk/20160105160709/http: //ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/methods/statistical-disclosure-control-methods/index.html

Geo-referencing	No geo-referencing issues
Availability of data	The data is available on the SIMD website www.gov.scot/SIMD. 2011 Census data is available on the Scottish Census website[19].

[19] http://www.scotlandscensus.gov.uk/

ANNEX A

Department for Work and Pensions data geo-referencing

A large scale exercise has been carried out by the Department for Work and Pensions to produce a single address for every individual at any point in time. This single address is based on the latest address that has been notified to the department in respect of any of the key benefits within the Work and Pensions Longitudinal Survey.

Geographic referencing was carried out by the DWP using the Customer Information System (CIS), Census Output Area (COA) and National Statistics Postcode Lookup (NSPL). In accordance with ONS geography policy, each record was allocated to a fixed geographical area, namely Census Output Area, and then all other higher geographies built up from this building block. Assigning a record to a 2011 COA is done in a couple of iterative stages:

Stage 1 – The Customer Information System history file contains historic COA spells for every claimant in the CIS by their encrypted National Insurance number. COAs are added to this data by matching with full postcodes. Where postcodes are incorrect, incomplete or missing, logical allocation takes place to find the COA that matches most closely to the available geographic information. This represents the most accurate source of geographical information for claimants.

Stage 2 – Higher level geographies were merged on to the National Statistics Postcode Lookup using 2011 COA. The NSPL hierarchy was designed by ONS.

ANNEX B

Geographic access domain methodology

Geographic Information Science and Analysis Team

Scottish Government

1. Introduction

The geographic access domain was added to SIMD in 2004 to model the ability of Scottish citizens to reach a number of key services. Geographic Access consists of two sub-domains; one-way public transport journeys and journeys made by private car. This accounts for both citizens who will be wholly dependent on public transport and car owners.

The domain measures aspects of access deprivation and is relevant to all people, as it is equally important to be able to access key local services in both rural and urban areas[20]. The development of the domain since 2004 is summarised in Table 1.1.

Table 1.1: Development of the Geographic Access Domain, 2004-2020

	SIMD 2004	***SIMD 2006, 2009, 2012, 2016, 2020***
Drive Times to…	• GP Surgeries • Fuel Stations • Post Offices • Supermarkets • Primary Schools	• GP Surgeries • Fuel Stations • Post Offices • Retail Centres • Primary Schools • Secondary Schools
Public Transport Times to…	N/A	• GP Surgeries • Post Offices • Retail Centres

For SIMD 2004, the Geographic Access domain was based purely on the time taken to drive to basic services. For SIMD 2006 onward secondary schools were added to the list of destinations. Supermarkets were removed in favour of retail centres, to capture areas with clustered multiple retail outlets such as high streets. In addition to changes in destinations, SIMD 2006 also introduced the analysis of access to basic

20 https://www.webarchive.org.uk/wayback/archive/20180515173724/http://www.gov.scot/Publications/2004/10/20089/45177

services by public transport and/or walking. The same journey types and destinations have been used for all SIMD analyses since 2006.

2. Data Sources

The data sources used to calculate the Geographic Access domain for SIMD 2016 and 2020 are summarised in Table 2.1, and discussed in further detail below.

Table 2.1: Data Sources of the Geographic Access Domain, 20012-2020

	Dataset	***SIMD 2016***		***SIMD 2020***	
		Source	***Records***	***Source***	***Records***
Origins	*Census Output Area Population Weighted Centroids*	*with either* NRS 2011 Census population (for weighting non-school destinations) *or* Primary / Secondary School Census population, SG 2013-15	46,351	*with either* NRS 2011 Census population (for weighting non-school destinations) *or* Primary / Secondary School Census population, SG 2016-18	46,351
Destinations	*GP Surgeries*	OS PointX (March 2015)	1,126	OS PointX (September 2018)	1,101
	Fuel Stations	OS PointX (March 2015)	889	OS PointX (September 2018)	938
	Post Offices	Post Office (2015) plus OS PointX (March 2015)	1,276	Post Office (October 2018) plus OS PointX (September 2018)	1,217
	Retail Centres	CACI Retail (2015)	464	CACI Retail (2018)	478
	Primary Schools	OS PointX (March 2015)	2,052	Scottish Government School Roll (October 2018)	2,037
	Secondary Schools	OS PointX (March 2015)	368	Scottish Government School Roll (October 2018)	416
Transport Network	*Road Network*	OS ITN April 2015	N/A	OS Highways October 2018	N/A
	Road Speed data	INRIX Road Speed Data (April 2015)	N/A	TRAFFICMASTER speed data (October 2018)	N/A
	Public Transport Network	Traveline National Dataset and Associate of Train Operating Companies data (October 2015)	N/A	Traveline National Dataset and Associate of Train Operating Companies data (October 2018)	N/A
Output	*Data Zones 2011*	SG Data Zones, 2011	6,976	SG Data Zones, 2011	6,976

Origins

Census Output Area centroids were used as origins to represent the distribution of the population throughout Scotland for both SIMD 2016 and SIMD 2020. Census Output Areas were created by the National Records of Scotland (NRS) following the 2011

Census. The master postcode centroid of an Output Area is the postcode grid reference that is nearest the average centre, weighted by the population count, of all the postcode grid references for a Census Output Area[21].

The Census 2011 population totals for each Output Area are used to weight the results when aggregating the travel time results from Output Area centroids to Data Zones for both SIMD 2016 and SIMD 2020. The primary and secondary schools calculations do not use the census population totals, but instead use school age population totals for three years; for SIMD 2020, this was for the years 2016 to 2018. These include students at primary or secondary stage from the Pupil Census from September each year, and are summed for each Output Area.

Destinations: GP Surgeries, Fuel Stations

For both SIMD 2016 and SIMD 2020, locations for GP surgeries and fuel stations were taken from the Ordnance Survey PointX product. PointX includes data for all of Scotland, plus a 10 kilometre buffer into England so that accessibility can be appropriately observed for those in Scotland who live near the border. PointX provides two sets of coordinates per point; one set representing the physical location of the building and a second set representing the nearest point on the Ordnance Survey Integrated Transport Network / Highways product, so that the points can be used in network analysis; the latter were the coordinates used for the SIMD accessibility analyses.

Destinations: Post Offices

For both SIMD 2016 and SIMD 2020, data were obtained from the Post Office, and included the name, location and weekly hours for all conventional branches and outreach services across Scotland and a 10km buffer into England. This was deemed to provide more complete coverage than the post offices included in the OS PointX product. Any Post Office service providing at least 6 hours or more per week was included in the analysis, including mobile post offices. Home delivery services were not included in the analysis as they cannot be represented by a fixed point.

In SIMD 2016 it was noted that the coordinates provided by the Post Office were not all spatially accurate. When checked against background mapping, it was found that some points were located off of the road network and sometimes at a distance from the true Post Office location. To correct this issue, the Post Offices were either matched with their record in the PointX dataset to find the coordinates on the road

21 https://spatialdata.gov.scot/geonetwork/srv/eng/catalog.search;jsessionid=647FF82D02E05C7962CE3E99C089E76B#/metadata/9d977a73-7884-4870-ae76-afccf8e6fae8

network, or manually checked and relocated to the nearest location on the road network.

Similarly, for the 2020 data, Post Offices were first matched to PointX to attach the coordinates on the road network. For those not included on the PointX dataset, some could be matched to the SIMD 2016 dataset, and the remaining Post Offices were manually plotted.

Destinations: Retail Centres

For both SIMD 2016 and SIMD 2020, retail centre locations were obtained from CACI Retail Footprint (2018) data. The Retail Footprint product defines the catchments of retail centres and is generated from credit and debit card data[22]. The retail centre locations are provided as coordinates, and the dataset includes a 10km buffer into England.

Destinations: Primary and Secondary Schools

Primary and secondary schools are the only two destination types which have a different source for SIMD 2020 from the one used in SIMD 2016. In 2016, the data for the Primary Schools and Secondary schools were sourced from the Ordnance Survey PointX product. However, it was noted by the Education Team in Scottish Government that, in PointX, schools with both a primary and secondary department are only included once, typically as a secondary school. Additionally, junior highs (schools with a primary department and a secondary department typically reaching only S3/S4) are mostly not included in the data supply.

In order to get a more accurate list of schools, the SIMD 2020 uses School Roll data from the Education Team within Scottish Government for all Scottish schools, supplemented with the Ordnance Survey PointX data for a 10km buffer into England. The data provided by the Education Team includes all primary and secondary Local Authority-funded schools, plus one school funded by the Scottish Government, which were open in October 2018. As per SIMD 2016, Independent and Special schools were removed from the analyses.

The change to the source data for the primary and secondary school locations has had a very small impact on the results; rather than the points being 'snapped' to the nearest road segment, they are located on the buildings, meaning the software includes a very short walk as part of the journey time (around 30 seconds) to get from the road to the building. In SIMD 2016 the school journeys ended on the road segments.

[22] http://www.caci.co.uk/products/product/retail-catchments

Road Network

For SIMD 2020, the Ordnance Survey Highways product was used to model Scotland's roads. This product is a re-branded version of the Ordnance Survey Integrated Transport Network, which was the used for SIMD 2016. Each segment of the road is assigned a road type (e.g. motorway, A-road, B-road etc.) and allocated an unique Topological Identifier (TOID), which can be used to join other data to the network (e.g. speed data, see below).

Road Speeds

Up to 2009, SIMD calculations were run using default road speed profiles from the software. From SIMD 2012 onward, observed road speeds have been used for the drive time calculations. The data used in SIMD 2020 are compiled by TRAFFICMASTER and distributed via Basemap. The data are derived from GPS units on a mixture of vehicle types, with average speeds recorded for different time periods. The time periods of collection which are relevant for the SIMD calculations are shown in Table 2.2.

Table 2.2: Data Collection Periods for TRAFFICMASTER 2018 Data

Peak	Off-Peak	Peak
07:00 to 09:00 Mon to Fri	10:00 to 16:00 Mon to Fri	16:00 to 19:00 Mon to Fri

The October 2018 speed data were applied to a 2017 version of the road network, therefore, some road segments from the October 2018 Highways product used in the SIMD analyses do not have a corresponding speed recorded in the TRAFFICMASTER dataset. For these segments, the average TRAFFICMASTER speed per road type for each of the time periods shown in Table 2.2 was used. For road types for which there are no TRAFFICMASTER data available, the TRACC Global Speed data is used for the appropriate time period, but adjusted to make it more realistic by reducing the speeds by 5 kph, in line with guidance from Basemap[23]. The two road types for which TRACC sets a Global Speed of 0 kph – (Restricted) Secondary Access Roads – were set to 10 kph, otherwise Output Area centroids which snap to this type of road would automatically be classed as inaccessible by TRACC, regardless of the availability of other nearby traversible road segments.

The speed data used in SIMD 2016 was from a different product, INRIX, but was also gathered from GPS units for the same peak morning and peak evening time periods, and an off-peak time period of 9am to 4pm. INRIX derives speeds for road segments

[23] https://www.basemap.co.uk/speed_dataset_tracc/

which are not covered by the GPS data based on the speeds of surrounding road segments.

The decision to use TRAFFICMASTER instead of INRIX for SIMD 2020 was based on advice from Basemap, who noted that TRAFFICMASTER is better suited to the OS Highways product within TRACC (the software used to calculate journey times).

Public Transport Network

For both SIMD 2016 and SIMD 2020, the public transport data were compiled from the Traveline National Dataset provided via Basemap. This is provided as a quarterly dataset including bus, rail, coach, ferry and tram data. The Scottish Government Ferry Routes dataset was used to identify a small number of ferry routes which should be disabled as they are tourist cruises (e.g. the Waverley Paddle Steamer), as well as the passenger routes which are disabled for the drive time calculations.

Output geography

For both SIMD 2016 and SIMD 2020, the journey times for Output Areas are weighted and applied to Data Zone boundaries from 2011.

3. Methodology

Accessibility Calculation

Basemap TRACC software is used to find the closest of each type of destination to every origin, and its travel time, by both driving and public transport. The travel time calculations must be performed for each Output Area centroid, then the results weighted and aggregated to Data Zone level to provide an average travel time for each Data Zone. Drive times were calculated for all six destination types, and public transport times were calculated for three destination types. Public transport times were not calculated for schools, because these tend to use transport for which there is no available national data, or fuel stations.

Key points of the methodology are as follows and the calculations are summarised in Figure 3.1:

- All calculations were performed for a Tuesday, as this was deemed to be the most representative weekday.
- As drive time speeds and public transport availability varies by time of day, travel times were calculated at three time windows to incorporate peak morning, peak evening, and off-peak periods, and outbound and inbound trips.

- Drive times for schools were calculated for only two time windows, one for a morning peak journey, and another for the late afternoon. This aims to better represent the regular 'school run' pattern.
- Journey times must start and finish within the four hour time periods (three hour time periods for schools).
- A maximum journey time of three hours (180 minutes) is imposed.
- Both drive times and public transport journeys can make use of ferries. For the drive time calculations, passenger ferries are disabled.

TRACC finds the number of nearest destinations stipulated by the user by performing a crow-flies distance calculation to each destination until it has found the allotted number. Then, for each of the closest destinations identified per Output Area, a travel time calculation is made and the shortest journey time is outputted. For this reason it is desirable to have more than one nearest destination, as the physically closest destination may be separated from the origin by a geographical barrier, such as open water, whilst the second (or third etc.) nearest destination may be reached directly by road in a shorter journey time. This parameter was increased for the drive times from two to five destinations per origin between SIMD 2016 and SIMD 2020, to ensure all closest destinations via the network were calculated. This parameter value for public transport journeys was ten in both SIMD 2016 and SIMD 2020.

To calculate the drive times, TRACC routes from each origin to the nearest road network segment, then through the road network to the destination using the quickest route. For SIMD, we also need to ensure it is possible for the routes driven to include ferry transport. In TRACC it is not possible to combine public transport and road network calculations in the *road network* mode, so the software has to be 'tricked' into performing the same sort of analysis using the *public transport and road network* mode. This mode normally calculates a travel time as though the sections of road network traversed to and from public transport nodes are done so on foot. By default, TRACC allocates a walking speed of 4.8 kph to these sections of the journey. However, it was possible to over-ride the default walk speeds in the TRACC projects to the TRAFFICMASTER road speed values for different times of day, so TRACC will walk through the network at vehicular speeds. This also over-rides the TRACC default of 0 kph on Motorways, so therefore means that this road type is traversable in the drive time calculations. Using this mode means that turn restrictions and one-way streets are ignored, which may result in some slightly faster journey times than are likely in reality; this impact will be consistent between SIMD 2016 and SIMD 2020 results, which followed the same methodology. For both the SIMD 2016 and SIMD 2020 drive time calculations, following advice from Basemap, the setting to normalise the road network[24] was used, which can reduce the accuracy of the results slightly, but was the necessary in order to run the volume of calculations required.

[24] https://www.basemap.co.uk/tip-of-the-month-normalise-road-network/

For public transport travel times for both SIMD 2016 and SIMD 2020, the *public transport and road network* mode is used which links each origin to the closest public transport stop, if available, at the 4.8 kph walking speed through the road network. This uses only the road network, not footpaths, and excludes walking on motorways. Using as many different modes of public transport as necessary, TRACC then routes each trip through the public transport network to reach the destination. Distances of up to 500 m were permitted to interchange between public transport services, and 2 km to reach the initial public transport stop. If the destination is within 2 km of the origin, the travel time to the nearest destination would be the walk time if this is quicker than the public transport alternatives.

Processing

Results were processed using the same Python code which was used in SIMD 2016. This first calculates an average drive time and public transport time for each Output Area. If fewer than two journeys per Output Area per journey type were possible, averages were not calculated for the Output Area as it was deemed inaccessible. This occurred for three reasons for drive time calculations: no ferries are available within the time window, the ferry required is passenger only or the journey takes longer than 180 minutes. For public transport, this occurs due to no public transport being possible within the time window, or the journey taking more than 180 minutes. Inaccessible drive times were adjusted to 190 minutes to represent the poor accessibility of the area. For inaccessible public transport calculations, the average was calculated as 2.5 times the equivalent drive time of the Output Area or, if the drive time had been imputed to 190 minutes, the public transport time was also imputed as 190 minutes.

To account for a wait time required to board before ferry journeys, the Traveline dataset occasionally builds an extra ten minutes into the timetabled journey time for ferries. Analysis of the data revealed this was not applied consistently to all ferry routes. Additionally, ten minutes was not seen as a realistic period of wait time. Therefore, for both SIMD 2016 and 2020, a wait time of 30 minutes for drive times and 20 minutes for public transport was applied if the journey utilised a ferry. This was done by identifying those ferries which have already had a wait time added and cross-referencing this with a list of Output Areas which should have wait times added due to their requirement to travel by ferry. Then the required period of time is added to the average travel time.

Following the addition of wait times, if the average journey time for the Output Area for drive times or public transport times was greater than 180 minutes, the journey time was adjusted to 190 minutes. This means that, as with SIMD 2016, no Output Area has an average drive time or public transport time greater than 190 minutes.

Output Area journey times were population weighted using 2011 census figures from NRS. For primary and secondary schools, a sum of three years (2016 to 2018) school

age population was used. For some Output Areas, there were no pupils in the Output Area across the three-year period. For both SIMD 2016 and SIMD 2020 this was altered so that a pupil population of 0.5 pupils was imputed to ensure that all Output Area results contribute to the Data Zone results.

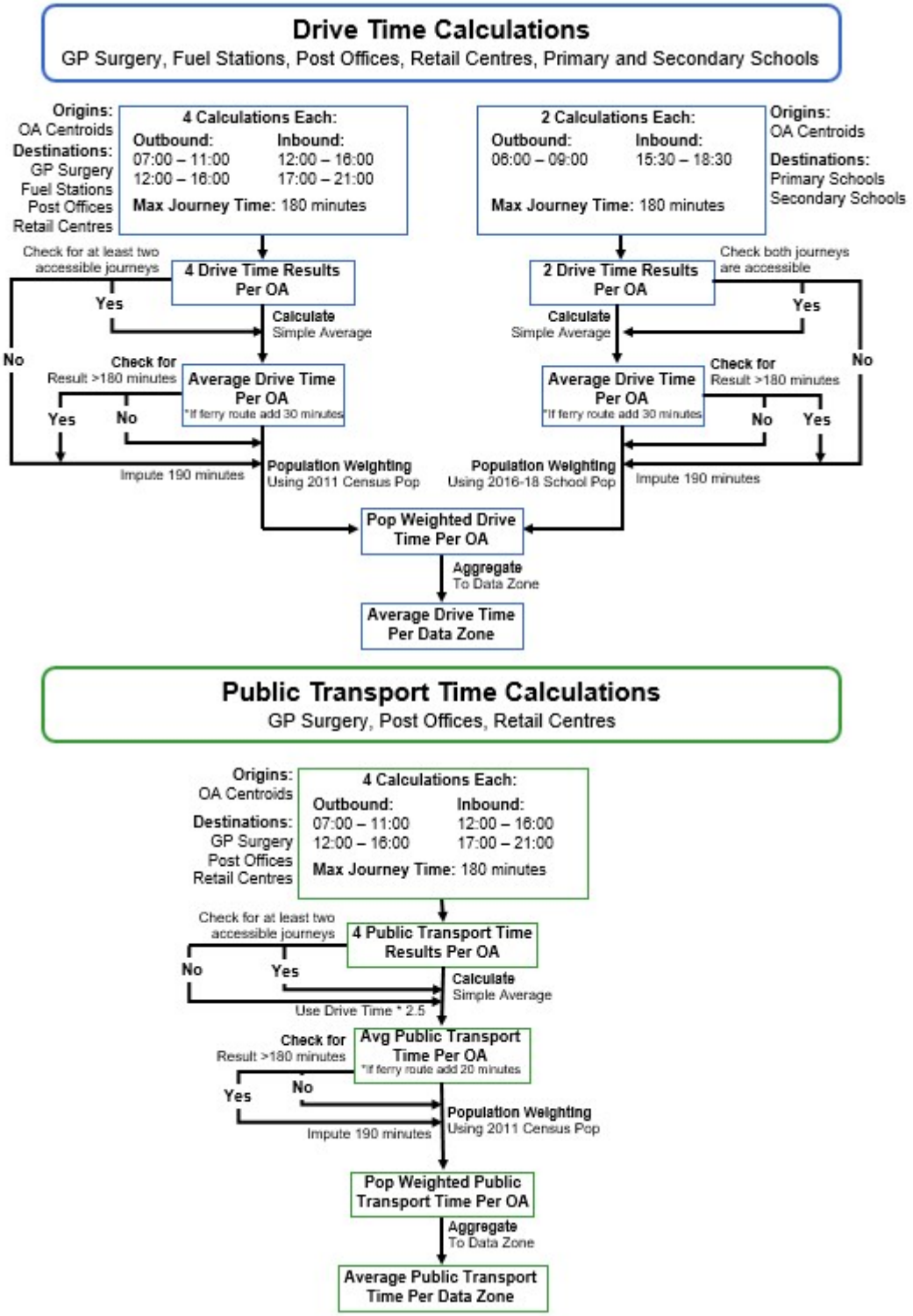


Figure 3.1: Summary of SIMD 2020 calculations and processing workflow

4. Output

Drive time averages in minutes, for each Data Zone, were provided for fuel stations, GP surgeries, post offices, retail centres, and primary and secondary schools. Average journey time by public transport in minutes, for each Data Zone, were also provided for GP surgeries, post offices and retail centres. These results were checked to verify that the travel times were reasonable and any changes in travel