Methodology for producing the Scotland Community Needs Index 2021

Introduction

In this paper we outline the methodological approach for developing a **Community Needs Index** for Scotland (SCNI). The SCNI has been constructed using the same methodology, domain structure and geographic unit of analysis as adopted in the 2021 update of the Community Needs Index for England (ECNI 2021); and uses the same component indicators where possible. In this paper, we summarise the steps involved in developing the SCNI.

The Indicators

The table below outlines the key socio-economic indicators which have been included in the SCNI. These have been grouped into domains and subdomains:

- Civic Assets: Measures of the presence of key community, civic, educational and cultural assets in and in close proximity to the area
- Connectedness: Measures of connectivity to key services, digital infrastructure, social isolation and strength of the local jobs market
- Active and engaged community: Measures concerning the levels of third sector civic and community activity and low levels of participation and engagement

The table provides an overview of each of these indicators with metadata detailing:

- Source (included URL)
- Timepoints the data is available for

- Geographical unit at which the data is published
- Relevance
- Notes associated with the indicator – including robustness issues to consider when incorporating the data

Indicator	Details	Source	Date	Granularity	Notes/Caveats
Civic Assets					
CA1: Density of community space assets	This is conceptualised as the number of community and civic assets inside the local area or within 1km of the local area boundary, divided by the number of people living in inside the local area or within 1km of the local area boundary. Rate is expressed per 100,000 population. The following assets are included: • Public / Village Hall / Other Community Facility • Youth Recreational / Social Club • Church Hall / Religious Meeting Place / Hall • Community Service Centre / Office • Place Of Worship	AddressBase https://www.ordnancesurvey.co.uk/business-government/products/addressbase	July 2021	Point Location	Details are not available on how accessible the assets are to the community.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CA2: Density of educational assets	This is conceptualised as the number of community and civic assets inside the local area or within 1km of the local area boundary, divided by the number of people living in inside the local area or within 1km of the local area boundary. Rate is expressed per 100,000 population The following assets are included: • College • Further Education • Higher Education • Children's Nursery / Crèche • First School • Infant School • Junior School • Middle School • Primary School • Secondary School • University • Special Needs Establishment. • Other Educational Establishment	AddressBase https://www.ordnancesurvey.co.uk/business-government/products/addressbase	July 2021	Point Location	Details are not available on how accessible the assets are to the community.
CA3a: Density of sport and leisure assets (address base)	This is conceptualised as the number of community and civic assets inside the local area or within 1km of the local area boundary, divided by the number of people living in inside the local area or within 1km of the local area boundary. Rate is expressed per 100,000 population. The following assets are included: • Public House / Bar / Nightclub • Activity / Leisure / Sports Centre • Skateboarding Facility • Recreational / Social Club(Bingo)	AddressBase https://www.ordnancesurvey.co.uk/business-government/products/addressbase	July 2021	Point Location	Details are not available on how accessible the assets are to the community. Some of the facilities identified will have a cost associated with access, which could potentially exclude those on lower incomes in the community.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CA4: Density of cultural assets	This is conceptualised as the number of community and civic assets inside or within 1km of the local area boundary divided by the number of people living in the inside or within 1km of the local area boundary. Rate is expressed per 100,000 population. The following assets are included: • Library • Reading Room • Museum/Gallery	AddressBase https://www.ordnancesurvey.co.uk/business-government/products/addressbase	July 2021	Point Location	Details are not available on how accessible the assets are to the community. Some of the museums will not be free to enter, which will exclude some sections of the community. Some of the libraries and reading rooms will not have open access.
CA5a: Green assets (density)	This is conceptualised as the number of community and civic assets inside or within 1km it divided by the number of people living in the inside or within 1km of the local area boundary. Rate is expressed per 100,000 population. The following assets are included: • Public Park / Garden • Public Open Space / Nature Reserve • Playground • Play Area• Paddling Pool• Picnic / Barbeque Site• Allotment• Playing Field • Recreation Ground	AddressBase https://www.ordnancesurvey.co.uk/business-government/products/addressbase	July 2021	Point Location	Details are not available on the accessibility of the asset from within the community. Some assets are not open- access to the whole community, e.g. allotments and some of the play areas/paddling pools. It is not possible to distinguish between these (though private parkland has been excluded). There is no information regarding the size or quality of the green space.
CA5b: Green assets (Area of public green space)	Area of public green space. Shows the average combined size of Parks, Public Gardens, or Playing Fields within 1,000 m radius (m2). Data is based on analysis of Ordnance Survey (OS) data on access to private gardens, public parks and playing fields in Great Britain.	Ordnance Survey https://www.ons.gov.uk/economy/environmentalaccounts/datasets/accesstogardensandpublicgreenspaceingreatbritain	April 2020	IZ	

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CA6: Retail assets	Number of retail premises in the local area or within 1km of the local area boundary) divided by the number of people living in the inside or within 1km of the local area boundary. The rate is expressed per 100,000 population. The following assets are included: Post Office, Market, Restaurant / Cafeteria, Shop / Showroom and Garden Centre	AddressBase https://www.ordnancesurvey.co.uk/business-government/products/addressbase	July 2021	Point Location	Does not take into account the size of the retail unit or how accessible it is to the local community. Excludes assets with negative community benefit
CA7: Community-owned assets	Community owned assets in divided by the number of people living in the inside or within 1km of the local area boundary. The rate is expressed per 100,000 population. Figures are compiled using data from Power to Change, the Community Land Trust Network, Co-operatives UK, Plunkett Foundation and Locality and Keep it in the Community.	Renaisi/ Plunkett Foundation/Locality	2021	Postcode	Some assets are geolocated based on the location of the organisation owning the assets rather than the assets itself, and some postcodes containing multiple assets are listed as single assets in the database.
Connectedness					

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CN1a: Travel time to key services by public transport/walk	Travel times in minutes to key services by public transport/walking and cycling. The following services are included: • GP • Shopping Facilities • Post Office These statistics are derived from the analysis of spatial data on public transport timetables; road, cycle and footpath networks; population and key local services.	Scottish Government https://statistics.gov.scot/resource?uri=http%3A%2F%2Fstatistics.gov.scot%2Fdata%2Ftravel-times	2015	DZ	Although the statistics are calculated to a high level of geographical detail, some assumptions and simplifications are necessary in the modelling (for example assigning the start point of journeys to a single point in each Output Area, road speeds, interchange times for public transport).
CN1b: Average distance to nearest Park, Public Garden, or Playing Field (m)	Average distance to the nearest park, public garden or playing field in meters. Data is based on analysis of Ordnance Survey (OS) data on access to private gardens, public parks and playing fields in Great Britain.	Ordnance Survey https://www.ons.gov.uk/economy/environmentalaccounts/datasets/accesstogardensandpublicgreenspaceingreatbritain	April 2020	IZ	

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CN2a: Jobs density in the Travel to Work Area	The number of jobs located in the area as a percentage of the working-age population in that area – this is to be used as a measure of economic opportunities locally. Data are taken from the Business Register and Employment Survey (BRES) of approximately 80,000 businesses, weighted to represent all sectors of the UK economy. The BRES definition of an employee is anyone aged 16 years or over at the time of the survey, whom the employer pays directly from its payroll(s) in return for carrying out a full-time or part-time job or for being on a training scheme. This indicator will be calculated at travel-to-work-area (TTWA) level rather than at community-geography level, to reflect the fact that people typically commute outside of their local area to work1. TTWAs are a geography created to approximate labour-market areas. In other words, they are designed to reflect self-contained areas in which most people both live and work. The current ONS criteria for defining TTWAs are that at least 75% of the area's resident workforce work in the area, and at least 75% of people who work in the area also live in the area. The area must also have an economically active population of at least 3,500.	Business Register and Employment Survey (BRES) https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=57	2019	TTWA	This measure does not take into account the quality of the job, whether they are full or part time, zero hours or temporary or permanent contract or how easily accessible the core of the travel to work area is from the specific community geography area.

1 More than half of those in employment travel more than 5km to work, with the average distance travelled to work across the England and Wales - 15km – Source: Census 2011 Distance travelled to work

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CN2b: Jobs density in the local area	The number of jobs located in the area as a percentage of the working-age population in that area – this is to be used as a measure of economic opportunities locally. Data are taken from the Business Register and Employment Survey (BRES) of approximately 80,000 businesses, weighted to represent all sectors of the UK economy. The BRES definition of an employee is anyone aged 16 years or over at the time of the survey, whom the employer pays directly from its payroll(s) in return for carrying out a full-time or part-time job or for being on a training scheme. This indicator will be calculated at based on the number of jobs inside or within 1km of the local area boundary to balance and ranked alongside the Jobs Density measure to get a weighted measure of local jobs and jobs in the wider labour market.	Business Register and Employment Survey (BRES) https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=57	2019	DZ	This measure does not take into account the quality of the job, whether they are full or part time, zero hours or temporary or permanent contract.
CN3: Households with no car	The proportion of households who do not have a car or van. Figures are based on responses to the 2011 Census car ownership question, which asks for information on the number of cars or vans owned or available for use by one or more members of a household. It includes company cars and vans available for private use. This is included to supplement the accessibility of key services and labour market indicators in this domain, to take account of the additional challenges in accessing services for those without access to private transport.	Census 2011 https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=621	2011	Output Area	The count of cars or vans in an area is based on details for private households only. Cars or vans used by residents of communal establishments are not counted.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CN4a: Broadband download speeds	Average broadband download line-speed (Mbit/s) for connections in the area.	OfCom	2020	Postcode	Due to variations in broadband performance over time, this data should not be regarded as a definitive and fixed view of the UK's fixed broadband infrastructure. However, the information provided here may be useful in identifying variations in broadband performance.
CN4b: Broadband upload speeds	Average broadband upload line-speed (Mbit/s) for connections in the area.	OfCom	2020	Postcode	Due to variations in broadband performance over time, this data should not be regarded as a definitive and fixed view of the UK's fixed broadband infrastructure. However, the information provided here may be useful in identifying variations in broadband performance.
CN5: Loneliness (People living alone)	Shows the proportion of households that comprise one person living alone (as a proportion of all households). Figures are self-reported and taken from the household composition questions in the 2011 census.	Census 2011 https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=605	2011	Output Area	This is included as a proxy measure of social isolation.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
CN5b: Loneliness (Loneliness Index – GP Prescriptions for Loneliness)	An outcome-based loneliness index using open prescription data. Open prescription data lists medicines, dressings and appliances prescribed by NHS primary care facilities, including General Practices (GPs), each month. Loneliness Index is created by using GP prescription data to find areas with above-average prescriptions for five conditions where loneliness has been shown to be a risk factor: Alzheimer's, depression, high blood pressure, anxiety and insomnia. An index was created for each condition by standardising the proportion of a practices prescriptions that were given for the condition relative to the levels in other practices (into z scores). The index for each condition had a value that was negative if prescribing was lower than typical and positive if it was greater than typical. The loneliness index is generated by summing together these standardised-scores for each condition.	Office for National Statistics' Data Science Campus /NHS /Red Cross https://github.com/matthewgthomas/loneliness/blob/master/README.md	2019	IZ	These data do not include any information about the person it was prescribed to and are averaged for a whole GP practice.
CN4c: Loneliness (Self-reported levels of loneliness)	People who have self-reported that they feel lonely most, almost all or all of the time.	Scottish Household Survey https://www.gov.scot/collections/scottish-household-survey/	2018	DZ	Data is available in the Scottish Household Survey by Local Authority and Scottish Index of Multiple Deprivation (SIMD) quintile and allocated to LSOAs based on their parent Local Authority and SIMD quintile score.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE1: Voter turnout at local elections	Valid votes turnout (%) at the most recent Local Council Elections	Electoral Commission https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-elections-and-referendums/european-parliamentary-elections/report-may-2019-european-parliamentary-elections-and-local-elections	2019	Ward	

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE2a: Civic participation (Self-reported measures of community and civic participation)	The Scottish Household Survey contains key indicators of volunteering and civic participation. We have explored datasets from the 2013, 2014, 2015, 2017, 2018 and 2019 iterations of the survey in order to produce a composite measure of participation. The following indicators are included (with survey years in brackets): • People who do no volunteering (formally or informally) (2018) • People who feel they can influence decisions locally (2013, 2014, 2015, 2017, 2018, 2019)	Scottish Household Survey https://www.gov.scot/collections/scottish-household-survey/	2013, 2014, 2015, 2017, 2018, 2019 and 2020	DZ	The Survey is published broken down by Local Authority and Scottish Index of Multiple Deprivation (SIMD) decile. Using a combination of these variables it is possible to apportion response rates to LSOA level allocating response rates (%) to each LSOA based on their SIMD decile and parent Local Authority. Indicators are then combined over multiple years to boost the sample size and improve the robustness of the estimates.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE2b: Strength of local social relationships	The Scottish Household Survey contains key indicators of strength of local social relationships. The following indicators are included: • This is a neighbourhood where people are kind to each other • This is a neighbourhood where most people can be trusted • There are welcoming places and opportunities to meet new people • There are places where people can meet up and socialise • This is a neighbourhood where people from different backgrounds get on well together • This is a neighbourhood where local people take action to help improve the neighbourhood	Scottish Household Survey https://www.gov.scot/collections/scottish-household-survey/	2018 and 2019	DZ	The Survey is published broken down by Local Authority and Scottish Index of Multiple Deprivation (SIMD) decile. Using a combination of these variables it is possible to apportion response rates to LSOA level allocating response rates (%) to each LSOA based on their SIMD decile and parent Local Authority. Indicators are then combined over multiple years to boost the sample size and improve the robustness of the estimates.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE3: Leisure and cultural participation	The Scottish Household Survey contains key indicators of cultural and sport participation. We have explored datasets from the 2013, 2014, 2015, 2017, 2018 and 2019 iterations of the survey in order to produce a composite measure of participation. The following indicators are included (with survey years in brackets): • Participated in cultural activities (excluding reading) over last 12 months – (2018, 2019) • Visiting outdoors at least once a week (2013, 2014, 2015, 2017, 2018, 2019) • Participation in sport incl walking at least once a week (2013, 2014, 2015, 2017, 2018, 2019)	Scottish Household Survey https://www.gov.scot/collections/scottish-household-survey/	2013, 2014, 2015, 2017, 2018 and 2019	DZ	Data are constructed from a survey with a small sample size. Data has been apportioned down to Output Area level using Output Area Classification group membership – (which groups together Output Areas based on their shared socio-economic characteristics) and Local Authority/WIMD 2019 quintile. Caution should be applied when interpreting these results at small-area level because of the small sample size of the survey. Four years of data were used to increase the size of the response rate.
AE4a Barriers to participation – Language Barriers	The proportion of the population (aged 3 and over) who cannot speak English or cannot speak English 'well'. The indicator is self-reported from the 2011 Census. The proficiency in English classification corresponds to the tick box response options on the census questionnaire. This is used as a measure of barriers to participation in key services	Census 2011	2011	Output Area	

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE4b Barriers to participation – Population churn: CDRC Residential Mobility Index	This index provides an estimate of the "churn" of the residential population in the UK – the proportion of households that have changed between the end of 2020 and 2019. The estimates were built from linking administrative and consumer data, including electoral registers, consumer registers and land registry house sale data. This data enables researchers to explore annual variations in neighbourhood change at a small area geography. Crucially it also enables researchers to focus on yearly data rather than relying on decadal census data to estimate change. It is even possible to observe trends that have occurred since the collection of the most recent census.	Consumer Data Research Centre (CDRC)	2019/2020	DZ	

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE5: Third sector organisations	Non-overlapping count of 1) Registered charities from Charity Base 2) Co-operative societies from Co-operatives UK 3) Charitable Incorporated Organisations, Community Interest Companies, PRI/LTD BY GUAR/NSC (Private, limited by guarantee, no share capital) and Registered Societies from Companies House 4) Co-operative societies, community benefit societies, and former industrial and provident societies from Financial Conduct Authority Figure is expressed as a rate per 100,000 population.	Charities Commission https://charitybase.uk/chc . Co-operatives UK https://www.uk.coop/uk , Companies House http://download.companieshouse.gov.uk/en_output.html , from Financial Conduct Authority https://mutuals.fca.org.uk/	2021	Postcode	This is based on the location of organisations rather than on their area of operations (some will have a global focus). Larger charities ae excluded from this measure. This indicator is included in this theme to capture the level of third sector activity in the local area. Organisations with an exclusively national or international focus have been excluded, to ensure only organisations with a local focus are included. Some organisations appear on multiple registers – duplicate records have been stripped so only unique records remain. This will exclude smaller companies not registered and exclude co-operatives, community benefit societies, associations, trusts and partnerships (of varying types)

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE6: National Lottery Community Fund	Combined total of grants made to local projects and organisations by the National Lottery Community Fund between 2004 and 2021 per 1,000 population (£). Figures are taken from data on grants made to projects and organisations in local areas in the UK by the Big Lottery Fund, from grants data published by Big Lottery in conjunction with the 360Giving initiative. Big Lottery used the 360Giving standard to produce a file of all the grants made in 2004-2021.	National Lottery (through 360 Giving) https://grantnav.threesixtygiving.org/	2004-2021	Ward level	Included in the active/engaged community theme to capture the level of third sector activity in the local area.

AE7: Grant funding per head from major grant funders	Combined grant funding from grant giving organisations whose data has been subject to the 360giving standard (per head of population). The following organisations are included: A B Charitable Trust, Access to Justice Foundation, Andrew Lloyd Webber Foundation, Barrow Cadbury Trust, CHK Foundation, Cabinet Office, Calouste Gulbenkian Foundation, UK Branch, Co-operative Group, Coop Foundation, Department for Business, Energy and Industrial Strategy, Department for Culture, Media and Sport, Department for Digital, Culture, Media & Sport, Department for Digital, Culture, Media and Sport, Department for Education, Department for Environment, Food and Rural Affairs, Department for International Development, Department for International Trade, Department for Transport, Department for Work and Pensions, Department of Health, Department of Health and Social Care, EsmÃ©e Fairbairn Foundation, Gatsby Charitable Foundation, HM Revenue & Customs, Hazelhurst Trust, Home Office, Imperial Health Charity, Indigo Trust, John Ellerman Foundation, John Moores Foundation, Joseph Levy Foundation, LGBT Consortium, LandAid Charitable Trust, Lloyd's Register Foundation, Lloyds Bank Foundation for England and Wales, London Marathon Charitable Trust, Masonic Charitable Foundation, Mercers' Charitable Foundation, Ministry for Housing, Communities and Local Government, Ministry of Defence, Ministry of Housing, Communities & Local Government, Ministry of Justice, National Churches Trust, National Emergencies Trust, Nationwide Foundation, Nesta, Nuffield Foundation, OVO Foundation, Paul Hamlyn Foundation, Pears Foundation, Power to Change Trust,	360 Giving Grant Nav data https://grantnav.threesixtygiving.org/	Up to 2021	Postcode level2	Data are based on the location of grant recipients rather than the location of their beneficiaries. This indicator is included in this theme to capture the level of third-sector activity in the local area. Grants above £1m excluded to ensure capturing local initiatives rather than national activity. Measure expanded to include all Grant Funders which have a nationwide focus (e.g. not focused in one region of the country only3) where geographic information supplied.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
	Rank Foundation, Road Safety Trust, Rothschild Foundation, Samworth Foundation, Sport England, Staples Trust, Tedworth Charitable Trust, The AIM Foundation, The Badur Foundation, The Bishop Radford Trust, The Blagrave Trust, The Clothworkers Foundation, The David & Elaine Potter Foundation, The Dulverton Trust, The Dunhill Medical Trust, The Fore, The Henry Smith Charity, The Joseph Rank Trust, The Michael And Betty Little Trust, The Pilgrim Trust, The Rayne Foundation, The Seafarers' Charity, The Segelman Trust, The Tudor Trust, Three Guineas Trust, True Colours Trust, Tuixen Foundation, Virgin Money Foundation, Vision Foundation, Wates Family Enterprise Trust, Wates Foundation, William Grant Foundation, Wolfson Foundation, Woodward Charitable Trust, ZING, the Trussell Trust				

2 Comic relief and Children in Need only supply references of Local Authority recipients. This data will be allocated to local area by apportioning

3 Grant givers with a specific area focus e.g. Community Foundations have been excluded to mitigate against reflecting the extent to which local grant givers have submitted data to GrantNav e.g. Not all Community Foundations have submitted data to Grant Nav and we want to guard against introducing systematic bias into the data by including data for some regions and excluding others.

Indicator	Details	Source	Date	Granularity	Notes/Caveats
AE8a: SME lending by banks	Total value of lending to SME businesses from key financial lenders (Barclays, CYBG, Lloyds Banking Group, HSBC, Nationwide Building Society, Royal Bank of Scotland and Santander UK in Great Britain).	UK Finance	June 2020	Postcode sector	Takes four quarters of lending data at postcode sector level. The data is modelled from postcode sector to Output Areas using a weighted lookup built from the numbers of shared postcodes between a postcode sector and Output Area in combination with the working age population per Output Area. Data is then aggregated to local area level to get total value of SME lending at local area level.
AE8b: Small businesses: Local Business Units with 0-4 employees	Small businesses: VAT registered local businesses with 0-4 employees per 10,000 population	Inter Departmental Business Register (IDBR)	2020	IZ	

Other indicators explored

Vacant Civic Assets

Address base contains a flag to identify unoccupied assets. We explored using this flag to identify community space assets, education assets, sport and leisure assets, cultural assets, retail assets and green assets that are classified as 'Unoccupied'. However, only 144 assets in the database were identified as 'unoccupied' in Scotland. This was an insufficient count to include in the analysis as the majority of IZs would have zero values.

Home Office Indicators of Integration.

The Indicators of Integration framework has been developed by the Home Office to identify the practical processes and changes that contribute to the integration of individuals and communities. It recognises that Integration is multi-dimensional and as such explores a range of indicators across 14 key domains that evidence suggests are of central importance to integration. The Social connections domain has the strongest overlap with the themes that the SCNI is intended to measure. There are three domains within the framework under the heading Social Connections:

- Social bonds: connections with others with a shared sense of identity
- Social bridges: connections with people of a different background
- Social links: connections with institutions, including local and central government services

We have explored the indicators under these sub-domains. All of the data is available from national surveys only so is insufficiently granular to include in the Index. However, we have cross-checked the indicators against relevant indicators in the Scottish Household Survey to ensure that these indicators are represented in the SCNI.

Scottish Social Attitudes Survey

The Scottish Social Attitudes Survey aims to facilitate the study of public opinion and inform the development of public policy in Scotland. It is a potential source of information on civic engagement and attitudes towards local government that could be used to measure engagement. However, the sample size is small 1,200-1,700 people – and does not have full national coverage - 93 postcode sectors are included, so it is insufficiently comprehensive and robust to be used in the SCNI.

National Records of Scotland

The National Records of Scotland (NRS) is a Non-Ministerial Department of the Scottish Government whose purpose is to collect, preserve and produce information about Scotland's people and history. The NRS contained potentially relevant data on migration and electoral statistics. However, the data on migration did not directly capture population churn and the electoral data was less granular than the data held by the electoral commission.

Community Ownership in Scotland

The Scottish Government produce an annual publication showing the extent of community ownership in Scotland. This data has been collected from a variety of sources: bodies providing funding to community groups to purchase assets, details of asset transfers from relevant public sector organisations and directly from community groups. The data has been collated and quality assured against the Land Register held by Registers of Scotland. We have validated this data against the more granular data sourced from Power to Change, the Community Land Trust Network, Co-operatives UK, Plunkett Foundation and Locality and Keep it in the Community.

Producing the SCNI

Step 1 Convert indicator data to IZ level

Each of the indicators in the SCNI is published at a different geographical level – however, in order to align with ECNI we intend to produce the SCNI at Intermediate Zone (IZ) level.

IZs are the preferred unit of measure because:

- They only change after every census, so they are more consistent over time. Even with changes made due to the census, about 95% will remain the same. They therefore represent a more stable geography than wards.
- They are generally all the same size (less than 5,000 people) but are sufficiently large enough that they are comparable to the average ward sizes.
- They are less politically linked (councillors or MPs) which was a criticism raised by a few stakeholders. That is, ward boundaries can be gerrymandered but IZs are not linked politically.
- They now have neighbourhood names (not just codes) so are more identifiable. They are also increasingly used to disseminate statistics releases – most notably the COVID-19 caseload - https://coronavirus.data.gov.uk/
- The list of 'left-behind' IZs can still be linked to wards and local authorities. With IZs, you might see more wards that are either fully or partially identified as 'left behind'.
- They nest directly with smaller statistical geographies such as Output Areas and Lower Data Zones (DZ) without requiring a best-fit lookup.

We therefore need to convert each of the underlying indicators to DZ geography for inclusion in the SCNI.

The table below outlines our approach to converting indicators to DZ level:

Geography	Indicators	Approach to conversion
Postcode/Point Location	Density of community space assets Density of educational assets Density of sport and leisure assets Density of cultural assets Green assets Retail assets Broadband speeds Grant funding per head from major grant funders Third sector organisations Community-owned assets	Use the ONS Postcode directory https://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-may-2021/about - to aggregate Postcode data to Output Area level. Use the Output Area to IZ Lookup table to aggregate data from Output Area to IZ level. Note for some of these areas a buffer zone is required around each of the IZs to allow for access to services outside of the IZ of residence.
Output Area	Households with no car People living alone Language Barriers	Use the Output Area to IZ Lookup table to aggregate data from Output Area to IZ level.
LSOA	Travel time to key services by public transport/walk Jobs Density in the local area Self-reported levels of loneliness Civic participation Strength of local social relationships Leisure and cultural participation CDRC Residential Mobility Index	Use the Output Area to DZ Lookup table to apportion data to Output Area. Use Output Area to IZ Lookup table to aggregate data from Output Area to IZ level.
IZ	Small businesses Green assets (Area of public green space) Average distance to nearest Park, Public Garden, or Playing Field Loneliness Index – GP Prescriptions for Loneliness	Use IZ boundaries directly
Electoral ward	Voter turnout at local elections National Lottery Community Fund	Apportion data from relevant ward to Output Area – using the appropriate Ward to Output Area look-up table – Output Area to Higher Area Index and apply Ward level scores to each Output Areas. Aggregate from Output Area to IZ using the Output Area to IZ Lookup table.

TTWA	Jobs density in the Travel to Work Area	Apply TTWA score to all the Output Areas in the TTWA using the LSOA to Travel to Work area lookup table. Aggregate from Output Area to IZ using the Output Area to IZ Lookup table.
Postcode Sector	SME lending by banks	Apply Postcode Sector score to all the Output Areas in the Postcode sector using the Output Area to Higher Area Index. Aggregate from Output Area to IZ using the Output Area to IZ Lookup table.

Step 2 Quality Assurance of the data

The next step is to comprehensively check the distributions of all indicators at IZ level to ensure that all indicators have passed the relevant fitness tests and are suitable for further analysis for the purpose of the SCNI. These tests include excluding indicators with high numbers of zeros or equal upper limit e.g. 100% values which would distort the Index.

Step 3 Applying shrinkage to improve the robustness of indicators

Where a rate or other measure of community need for an IZ is based on small numbers, the resulting estimate may be unreliable, with an unacceptably high standard error. The technique of shrinkage estimation is used to 'borrow strength' from larger areas to increase the reliability of small area data; the impact of shrinkage will tend to move an IZ's score towards that of their parent higher-level area. Shrinkage moderates the levels of unreliability in the dataset and reduced the impact of chance fluctuations from year to year. Such scores occur most commonly where numbers are small at IZ level and the event is thus relatively rare. This may be the case for the indicator as a whole or only for particular IZs. In shrinkage estimation the score for a small area is estimated as a weighted combination of that small area's score and the mean value for a larger area from which the smaller areas within the larger area borrow strength. The 2020 Local Authorities will be used as the larger area (this was the larger area used in the Scottish Indices of Deprivation shrinkage calculations). IZs within a single Local Authority share issues relating to local governance. To a certain extent, they may also share issues relating to labour market sub-climates. Shrinkage will be applied to all indicators with the exception of those published at Travel to Work Area (see table in Step 1 above).

Step 4 Ensuring that all indicators are "pointing in the same direction"

In order to combine the indicators into domains, it is necessary for each of the indicators to be orientated in the same direction. However, for some of the indicators included in this measure, a *high* value indicates *low* levels of need on the SCNI – for example an area with a high levels of grant funding would be

measured as having low levels of need. By contrast, for other indicators, a high score denotes high levels of need – for example areas with high travel times to key services. It is necessary therefore to 'reverse the polarity' for some scores to ensure that a high value is negative for all indicators – so they can be consistently combined.

Step 5 Producing composite indicators

A small subset of the indicators will be amalgamated to provide composite indicators before combining with the other indicators to create domain scores.

The following indicators will be grouped together:

Original indicators	Combined indicator
• Density of Green Assets • Area of public green space	Green assets
• Jobs Density in the Travel to Work Area • Jobs Density in the Local Area	Jobs Density
• Broadband upload speed • Broadband download speed	Broadband
• Travel time to GP • Travel time to Shopping Facilities • Travel time to Post Office • Average distance to nearest Park, Public Garden, or Playing Field	Physical connectedness
• People living alone • GP Prescriptions for Loneliness • Self-reported levels of loneliness	Social isolation[4]
• Population churn: CDRC Residential Mobility Index • Language Barriers	Barriers to participation
• Self-reported measures of community and civic participation • Strength of local social relationships	Civic participation
• Small businesses: Local Business Units with 0-4 employees • Small businesses: SME lending by banks	Small businesses

[4] Factor analysis is used to weight this indicator (see explanation below)

Before combining each of the individual indicators to produce an overall composite indicator, the indicators will first have shrinkage applied (to reduce any standard errors associated with small numbers), the indicators will then be standardised (by ranking and transforming to a normal distribution) – as each of the composite indicators are on a different scale (step 6 below describes the standardisation process in more detail). Note: Where there are more than two component indicators in a composite indicator – indicators will be checked for positive correlation and provided indicators are positively correlated, the weights of each component indicator will be determined using factor analysis (see step 8 below); where there are two indicators in a composite indicator, each indicator will be assigned an equal weight of 0.5.

Step 6 Standardisation

When combining measures, it is important to ensure that indicator scores are comparable and that the weighting of domains is not distorted by the fact that some of the indicators may have very different distributions. The indicators in the SCNI are based on different metrics and each indicator in the Index needs to be standardised to ensure that each indicator has a common distribution, so that indicators can be combined, without one indicator dominating due to a much larger distribution. Indicators will be standardised by ranking each of the indicators and then transforming to a normal distribution.

Step 7 Creating subdomains

The *Connectedness* and *Active and engaged community* domains will be split into subdomains. The principal reason for doing this is to reflect the character of the domains - as both sets of domains contain two conceptually distinct subsets of indicators.

The *Connectedness* domain explores connectivity both in terms of, access to services on the one hand and wider measures of connectivity on the other hand - such as access to transport, digital connectivity and isolation, which do not necessarily have strong associations with the more physical concepts of connectivity. We therefore propose grouping the domain into two subdomains:

Subdomain	Indicators
Physical connectivity	Physical connectedness Jobs Density
Wider connectivity	Households with no car Broadband speeds

	Social isolation

The Active and engaged community domain consists of measures concerning self-reported participation and engagement on the one hand, alongside measures of the strength of the community sector. Again, it makes conceptual sense to group these into separate subdomains as follows:

Subdomain	Indicators
Civic participation	Voter turnout at local elections Self-reported measures of community and civic participation Cultural participation Barriers to participation
Civic activity	Third sector organisations per head Big Lottery funding per head Grant funding per head from major grant funders Small businesses

An additional advantage of grouping indicators into subdomains is that it makes it possible to apply empirical weights to the indicators using factor analysis (explored in the weighting section below) – one of the key blockers of running factor analysis on the domains as a whole is that, within the Connectedness and Active and Engaged domains, the respective underlying indicators do not have a close association/a common underlying factor which explains their distribution. Grouping the indicators into smaller subdomains where indicators share the same factors enables factor analysis to be run to apply weightings to the majority of indicators.

Step 8 Weighting indicators

A statistical technique called Maximum Likelihood Factor Analysis is used to determine the weights of the indicators within each domain/subdomain. Factor analysis works most effectively where there is a single overwhelming factor which explains the performance on a set of indicators within a domain and where indicators within a domain exert an influence on one another. The outcome of applying factor analysis is that not all indicators in the domain will have equal weights, with the weights affected by the extent to which each of the indicators within a domain measure the underlying aspect that the domain is trying to capture. A key advantage of using factor analysis, is that it takes into account 'double-counting' within domains. We have run correlation analysis on the SCNI to determine the associations between indicators within each domain/subdomain. Following the outcome of this research we have applied factor

analysis to the *Civic assets* domain, the *Physical connectivity* subdomain, the *Civic activity* subdomain and the *Civic participation* subdomain of the SCNI. For the *Wider connectivity* subdomain each of the indicators will be assigned an equal weighting as each of the component indicators do not have close associations.

Step 9 Combining indicators to form subdomains and domains

The weighted and standardised indicators have then been combined to form subdomain scores (in the case of indicators in the *Active/engaged community* and *Connectedness* domains) and domain scores (in the case of the Civic assets domain – which does not contain any subdomains). The combination process involves summing each of the weighted indicator scores (the standardised indicator scores * weight) together for all the indicators within a domain/subdomain.

The subdomains are then standardised (using the exponential transformation method outlined in step 10 below) and then added together to form domain scores.

Step 10 Standardising domains

The three domain scores are then combined to produce the overall Scotland Community Needs Index (SCNI).

However, each of the domains are on a different scale to one another, two will be produced from combined subdomain scores, while the *Civic assets* domain will be produced from combined weighted indicators. It is therefore necessary to standardise the domain scores before combining. As with the 2021 English Community Needs Index, the method of standardisation that has been adopted is to transform the domains to a specified **exponential distribution** using an Exponential Transformation function. The exponentially transformed subdomain/domain scores are then combined to form an overall 'community need' measure at IZ level.

The Exponential Transformation method of standardisation differs from the normal distribution method as it gives more emphasis on the most deprived end of the distribution and so facilitates identification of the areas with the highest levels of need. This was the method of standardisation applied in the Indices of Deprivation in order to control cancellation effects e.g. high levels of deprivation in one domain are not completely cancelled out by low levels of deprivation in a different domain and ensures that areas that perform particularly badly on a particular aspect of community need are moved closer to the high end of the community need spectrum even when they show positive outcomes on other indicators.

Step 11 Weighting domains

The final stage for producing the Community Needs Index is to assign weights to the three domains that have been created – to apply to the domain scores before importing. It is important to note that all potential combinations of domains involve weights. If, after standardisation, the domains are simply added together, this gives each domain an equal weight. Our approach is for the weights to be explicit and based on clear criteria. We are committed to a procedure of combining Domain Indices in such a way that the weighting of the indices is explicit. Part of this commitment to transparent weights involves the standardisation of the Domain Indices as outlined above. This ensures that the domains can be combined without 'hidden' weights. Having standardised the domains, we are then able to choose explicit weights. We have adopted the same approach to weighting domain as agreed for the English CNI 2021.

Step 12 Creating a Scotland Community Needs Index (SCNI)

Once the preferred approach to producing domain weights is agreed, the domains can be combined to produce the overall Scotland Community Needs Index (SCNI). The combination process involves summing each of the weighted standardised domain scores (the exponentially transformed domain scores * weight) together to produce an overall SCNI score. This score is then ranked across all IZs in Scotland in order to combine with the SIMD score (step 13 below) to identify 'left-behind' areas.

Step 13 Identifying 'left-behind' areas

'Left-behind' areas are conceptualised as IZs in Scotland, which have high levels of need on both the SCNI and the Scotland Index of Multiple Deprivation 2020. Stages 1-12 are concerned with producing an *IZ* level Scotland Community Needs Index. The next step is to match this data against the Scotland Index of Multiple Deprivation 2020 – in order that both Indices can be analysed together to identify left-behind areas.

In line with the approach taken in the England CNI 2021, we have identified areas as 'left-behind' if they rank among the 10% of IZs in Scotland with the highest levels of Community Need and contain a DZ ranked among the 10% most deprived areas on the SIMD 2020.

Appendix A: Factor Analysis methodology

Factor analysis is used as a method for combining indicators, by finding appropriate weights for combining indicators into a single score based on the inter-correlations between all the indicators.

Factor analysis is only used in domains where 'latent variables' are hypothesised to exist and where the indicator variables are 'effect indicators', i.e. indicators that are influenced by the latent variable.

There are many candidates in terms of types of factor analysis. Two of the main contenders are maximum likelihood factor analysis (as used in the current and previous versions of the Indices of Deprivation) and Principal Components Analysis. The distinction between maximum likelihood factor analysis and Principal Components Analysis is a technical one. In brief, the assumptions underpinning Principal Components Analysis are that the indicators going into the analysis are perfectly reliable and measured without error. Maximum likelihood factor analysis requires no such assumption.

The process of combining indicators using factor analysis comprised three stages:

All indicators were converted to the standard normal distribution.

The standardised scores were factor analysed (using the Maximum Likelihood method), deriving a set of weights.

The indicators were then combined using these weights.

Appendix B: Exponential transformation

In order to combine the domains into an overall measure of need, the domain scores first need to be standardised. Any standardisation and transformation should meet the following criteria:

- Standard distribution. It must ensure that each domain has a common distribution, so that domains can be combined, without one domain dominating due to a much larger distribution.
- Identify areas of need. It must facilitate the easy identification of the areas with highest levels of need.
- Scale independent. It must not be scale dependent (in other words confuse population size with level of need).

One possible standardization approach involves each of the domain scores being ranked, and then the ranks are transformed to an exponential distribution. The exponential distribution has a number of properties that satisfy the criteria above.

Standard distribution

The exponential distribution transforms each domain so that they each have a common distribution, the same range and identical maximum / minimum values. The process starts by ranking the scores in each domain to standardise the domain scores (from 1 for the lowest need to 1,279 for the highest need), before applying the exponential transformation procedure to create a standardised domain score ranging from 0 (lowest need) to 100 (highest need).

Cancellation

The exponential transformation procedure gives control over the extent to which lack of need in one domain cancels or compensates for high need in another domain. It allows precise regulation, although not elimination, of these cancellation effects. The scaling constant (23) used produces roughly 10 per cent cancellation. This means that in the extreme case, an IZ which was ranked most deprived on one domain but least deprived on another would overall be ranked at the 90th percentile in terms of levels of need. This compares to the 50th percentile if the untransformed ranks or a normal distribution had been used instead.

Identify deprived areas

The exponential transformation effectively spreads out that part of the distribution in which there is most interest – that is the 'tail' which contains the areas with the highest levels of need in each domain. The scaling constant ensures that the most deprived 10 per cent of areas cover 50 per cent of the distribution of scores (in other words, scores between 50 and 100 after exponential transformation).

Scale independent

The transformation is not affected by the size of the IZ's population.

The exponential transformation calculation

The transformation used is as follows:

> For any IZ, denote its rank on the domain R, scaled to the range [0,1]. R=1/N for the least deprived and R=N/N (in other words R=1) for the most deprived, where N=the number of IZs in Scotland.
>
> The transformed domain score X is given by:
>
> $$X = -23 \ln\left(1 - R\left(1 - \exp^{-100/23}\right)\right)$$
>
> where 'ln' denotes natural logarithm and 'exp' the exponential or antilog transformation

Appendix C: Shrinkage

Improving the reliability of small area data values using shrinkage estimation

The shrinkage technique is designed to deal with the problems associated with small numbers in an IZ. In some areas – particularly where the at-risk population is small – data may be 'unreliable', that is more likely to be affected by sampling and other sources of error.

The technique of shrinkage estimation (in other words empirical Bayesian estimation) is used to 'borrow strength' from larger areas to avoid creating unreliable small area data. Shrinkage estimation involves moving IZ scores towards another more robust score, often relating to a higher geographical level. All IZ scores will move somewhat through shrinkage, but those with large standard errors (in other words the most 'unreliable' scores) will tend to move the most. The IZ score may be moved towards a 'higher need' or 'lower need' score through shrinkage estimation. Without shrinkage, some IZs would have scores which do not reliably describe the community need in the area due to chance fluctuations from year to year.

It could be argued that shrinkage estimation is inappropriate for administrative data which are, in effect, a census. This is not correct. The problem exists not only where data are derived from samples but also where scans of administrative data effectively mean that an entire census of a particular group is being considered. This is because such censuses can be regarded as samples from 'super-populations', which one could consider to be samples in time. All the data from administrative sources and the 2011 Census are treated as samples from a super-population in this way, and the shrinkage technique was applied to indicators which use this data. The exceptions are the indicators supplied at Local Authority District level.

Selecting the larger areas from which unreliable small area data can borrow strength

The principle for selecting the larger area should be that the IZs within them share characteristics. In the current shrinkage methodology, Local Authority Districts are used. The IZs within a single district share issues relating to local governance and possibly to economic sub-climates. To a certain extent, they may also share issues relating to labour market sub-climates.

The shrinkage calculation

The actual mechanism of the shrinkage procedure is to estimate deprivation in a particular IZ using a weighted combination of (a) data from the IZ, and (b) data from another more robust score (in the case of the Indices, this is the Local Authority District score). The weight attempts to increase the efficiency of the estimation, while not increasing its bias. All IZ scores are adjusted to some degree through the shrinkage process, but the magnitude of the adjustment will be greatest for areas with the least reliable scores. The amount of movement depends on both the size of the standard error and the amount of heterogeneity amongst the IZs in a Local Authority District.

The 'shrunk' estimate of an IZ level proportion (or ratio) is a weighted average of the two 'raw' proportions for the IZ and for the corresponding District. The weights used are determined by the relative magnitudes of within-Ward and between-Ward variability.

If the rate for a particular indicator in IZ j is r_j events out of a population of n_j, the empirical logit for each IZ is:

$$m_j = \log\left[\frac{(r_j + 0.5)}{(n_j - r_j + 0.5)}\right]$$

whose estimated standard error s_j is the square root of:

$$s_j^{\,2} = \frac{(n_j + 1)(n_j + 2)}{n_j(r_j + 1)(n_j - r_j + 1)}$$

The corresponding counts r out of n for the district in which IZ j lies gives the district-level logit:

$$M = \log\left[\frac{(r + 0.5)}{(n - r + 0.5)}\right]$$

The 'shrunk' IZ level logit is then the weighted average:

$$m_j^* = w_j m_j + (1 - w_j) M$$

where w_j is the weight given to the 'raw' IZ-j data and $(1-w_j)$ the weight given to the overall rate for the district. The formula used to determine w_j is:

$$w_j = \frac{1/s_j^2}{1/s_j^2 + 1/t^2}$$

where t^2 is the inter-Ward variance for the k IZs in the district, calculated as:

$$t^2 = \frac{1}{k-1} \sum_{j=1}^{k} (m_j - M)^2$$

Thus large IZs, where precision $1/s_2^j$ is relatively large, have weight w_j close to 1 and so shrinkage has little effect. The shrinkage effect is greatest for small IZs in relatively homogeneous districts.

The final step is to back-transform the shrunk logit m_j^* using the 'anti-logit', to obtain the shrunk IZ level proportion for each IZ:

$$z_j = \frac{\exp(m_j^*)}{1 + \exp(m_j^*)}$$