Scotland's Census
Shaping our future
A' dealbhadh ar n-àm ri teachd

Digital Exclusion Index for Scotland's Census 2022

External Methodology Assurance Panel, July 2020

National Records of Scotland

Table of Contents

1. Summary

A key objective for Scotland's Census 2022 is for people to complete their census questionnaires online when possible. However, not every household in Scotland is fully able or willing to complete the census online.

To be able to support these groups of people as much as possible it is necessary to establish where digitally excluded people live and what proportion of each area they make up. The Digital Exclusion Index ranks areas by how digitally excluded the people living in them are likely to be. This will allow resources to help people to take part in the Census to be directed to the right areas.

2. Executive Summary

The term "digitally excluded" refers to people who are not capable of using the internet, either because they lack digital skills or confidence in using a computer and online services and/or they do not have access to the internet[1]. Non-use of the internet could be due to many factors including:

- lack of digital skills to safely and competently use online services
- lack of confidence or interest in using a computer or online services
- no or slow access to an internet connection
- no access to devices or assistive technologies necessary to use the internet.

Digitally excluded people may need assistance of some form to respond to the census online or may need support to request paper questionnaires. Digitally reluctant people (who have access to the internet, and the skills to use computers and the internet, but prefer not to do so) are not covered by this definition.

In order to best provide assistance we need to determine the characteristics of digitally excluded people and which areas of Scotland they live in.

To aid census planning and operations Scotland is divided into approximately 9000 planning areas (PAs), each of which contains around 200 to 400 households. The Digital Exclusion Index is a ranked list of planning areas, based on how digitally excluded the households within them are thought to be. It is segmented into five categories in order to allow interventions to be targeted to different groups.

In order to create the Digital Exclusion Index we have:

- Researched factors underlying digital exclusion.
- Researched demographic data that can be used as a proxy for underlying variables.
- Assessed available data sources to map the geographic distribution of demographic factors across Scotland.

[1] ONS, 'Exploring the digital divide' https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04

- Used Scottish Household Survey data to link demographic factors with internet non-use.
- Created a model that uses demographic factors to predict internet non-use.
- Used national data sources to assign digital exclusion scores to every planning area in Scotland.
- Ranked all planning areas and segmented the ranked list into five categories.
- Tested the index and segmentation by analysing NRS census rehearsal 2019 data. We used this analysis to improve the methodology behind the index and the segmenting of the index into groups.

3. Introduction

During Scotland's Census 2022, resources will be directed towards enabling and encouraging the majority of census returns to be submitted online. However, not everyone is able to use a computer or the internet, and some people may struggle with filling in and submitting an online census questionnaire.

For Scotland's Census 2022, the Digital Exclusion Index (DEI) will be used to inform public assistance planning, including: predicting support hub use, predicting postal paper request volumes and predicting field force doorstep paper handout volumes. The DEI will also help inform discussions around the efficacy of delivering paper questionnaires as first contact to areas most affected by digital exclusion.

4. 2011 Census and 2022 Census

A Digital Exclusion Index (DEI) was not used in Scotland's Census 2011. In 2011, paper questionnaires were sent to every household, although an online census questionnaire was available.

Since the last census in 2011 digital services have become more common, in line with internet connectivity becoming more widespread. In 2019, Scottish Household Survey (SHS) estimates suggested that 88% of Scottish households have the option to have a fixed broadband connection to the internet[2]. Due to this expansion of connectivity, the majority of households in Scotland are expected to be able to complete their census questionnaire online in 2022.

5. ONS and NISRA approach to Digital Exclusion

In England and Wales, the Office for National Statistics (ONS) is using an index called the Hard to Count Index[3] that combines factors linked to low engagement with the Census ("willingness" in their documentation) with a digital exclusion component

[2] Scottish Household Survey 2019: annual report https://www.gov.scot/publications/scottish-household-survey-2019-annual-report/pages/8/
[3] ONS Hard to Count index for the 2021 Census. https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2020/07/EAP102-Hard-to-Count-index-for-the-2021-Census.docx

based on DVLA online and paper licence application data[4]. The Northern Ireland Statistics and Research Agency (NISRA) is sending paper to some households based on digital access[5].

The NRS Hard to Count index, addressed in a separate paper[6], is separate from the NRS Digital Exclusion Index described in this paper. NRS has separated the two because there are some interventions that are being considered that may specifically target digital excluded populations who may have relatively high engagement with the census, such as people over the age of 75. Conversely, there may be some groups who are highly digitally skilled but have in the past had low engagement with census, such as university students. Separating the indices may allow us to better target interventions to these groups.

6. Estimating Digital Exclusion: Underlying factors

6.1 Previous research

The common self-reported factors[7] that are associated with digital exclusion are:

- **Skills** and the confidence to use them
- **Access** to infrastructure
- **Cost** of equipment and service
- **Motivation/Interest** that makes gaining digital skills relevant and important.

In order to describe digital exclusion more directly, we looked at demographic factors associated with digital exclusion in various studies. Access is related to infrastructure rather than demography and will be addressed separately.

6.2 Digital Exclusion due to lack of internet access

The DEI initially included a variable on whether users have physical access to the internet (broadband connection/ smart phone/ library). This was measured using data on broadband speed connectivity in a geographic area. The proportion of households within planning areas that cannot receive internet speeds greater than 10Mbps was calculated from OFCOM data at postcode level and aggregated to planning area level, creating a Digital Access Map. However, later testing on data from the 2019 NRS Census Rehearsal indicated that this variable did not improve the ability of the model to predict key digital exclusion measures, and so it was

[4] ONS Update on methodology for the Digital domain of the Hard to Count index for the 2021 Census https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2020/07/EAP124-Hard-to-Count-index-methedology.pdf

[5] NISRA Key elements of the 2021 Census https://www.nisra.gov.uk/statistics/2021-census-proposals-document/key-elements-2021-census#toc-0

[6] NRS https://www.scotlandscensus.gov.uk/documents/Scotlands_Census_2022_-_Operational_Statistics_-_Hard_to_Count_EMAPs_paper_corrected(1).pdf

[7] E. J. Helsper and B. C. Reisdorf, "The emergence of a "digital underclass" in Great Britain and Sweden: Changing reasons for digital exclusion," *New Media & Society,* vol. 19, pp. 1253-1270, 2017.

removed from the index. Detail on the development of the Digital Access Map has therefore not been included in this report. See section 11 for further information.

7. Estimating Digital Exclusion: Demographic data

This section considers the demographic factors discovered in other studies to be linked to digital exclusion, and which of those variables are available in recent, high-quality, small-geography datasets.

Research from the ONS in 2019 suggests that digital exclusion is declining over time as increasing proportions of the population report being internet users[8], but shows there are still correlations between internet use and age, gender and ethnicity. Digital exclusion disproportionately affects low-income groups[9], people in older age groups[10] and the more vulnerable or marginalised communities in our society[11]. This suggests a strong correlation between digital exclusion and social exclusion.

We first consider what demographic data is available at low geographic levels, and select characteristics that are a) linked to Digital Exclusion and b) for which there is reliable, recent, small geography-level data for Scotland.

We then use this data to create a predictive tool which provides an indicative measure of internet users/ non-users for each area of Scotland. This tool can then be used to identify the areas facing high levels of digital exclusion.

7.1 Demographic factors considered

7.1.1 Data Zones

Data Zones are subdivisions of council areas in Scotland, they contain between 500 and 1,000 people and where possible boundaries have been constructed to respect natural communities, have regular shapes and contain households with similar social characteristics.

A range of Scottish Government statistical outputs are produced for Data Zones, including for income and age distributions.

7.1.2 Data Sources

The information produced by the Scottish Government at Data Zone level includes population distributions of age, level of education, income, employment and

[8] Internet use in the UK annual estimates by age, sex, disability and geographical location. Office for National Statistics 2019, based on figures from Labour Force Survey. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019

[9] The Digital Divide: What does the research tell us? Cambridge Centre for Housing and Planning Research, 2020 https://www.cchpr.landecon.cam.ac.uk/Research/Start-Year/2017/building_better_opportunities_new_horizons/digital_divide_research

[10] Digital Inclusion Evidence Review 2018. Age UK 2018 https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/age_uk_digital_inclusion_evidence_review_2018.pdf

[11] Exploring the UK's digital Divide. Office for National Statistics, 2019 https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04

disability. Each set of data is collected, and published from various sources with varying regularity including the 2011 Census, quarterly surveys and administrative sources such as tax returns. These and other potential predictors of digital exclusion available from other data sources are listed below in Table 1.

Age, income, and education have the most reliable, useful and available datasets.

Table 1: Factors considered as possible predictors of digital exclusion

Name	Data availability and suitability
Age	Mid-year Small Area Population Estimates for 2011 Data Zones provide annual data on age distributions at small geography level.
Income	The annual Scottish Government Local Level Distribution of Household Income Estimates gives the proportion of households in a Data Zone with weekly income under set intervals.
Education	Data on qualifications has been taken from the 2011 census bulk output tables. This dataset details the number of people aged 16 or over by the level of their highest educational or vocational qualification.
Disability	Disability information is taken from the Scottish Survey Core Questions. This is the highest quality data on self-reported disability by Data Zone but the statistic is measured on the criteria "Limiting Long Term Physical Or Mental Health Condition". There is a potentially confounding link between this metric and age-related disease.
Employment	The latest employment figures, supplied by the Labour Force Survey, are only available at local authority level. To have Data Zone level information would mean using the Census Outputs from 2011. Since employment is very dynamic these figures are likely to have changed since 2011.
Ethnicity	As ethnic minority groups are a relatively small proportion of the population of Scotland, particularly outside urban centres, it is unlikely that ethnicity will be a useful predictor of digital exclusion across much of Scotland. Ethnicity data is available in 2011 Census data.
Persons living alone	Household structure data is provided in the 2011 Census, but is likely to have shifted in geographic distribution over time
Sex	While there is some evidence that sex is related to internet use at an individual level in some age groups[12], when sex data is geographically aggregated it is likely that most areas will have close to equal numbers of men and women, as mixed-sex households and neighbourhoods are common. It is unlikely that inclusion of this variable would greatly influence the profile of the planning area.

[12] ONS, 2019: Internet use in the UK annual estimates by age, sex, disability and geographical location.
https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019#still-a-difference-in-internet-use-between-men-and-women-in-older-age-groups

8. Estimating Digital Exclusion: Internet use data from Scottish Household Survey

We used the results of the 2019 Scottish Household Survey (SHS) to measure the proportions of internet users and non-users in different areas. We then examined the correlation between measures of age, income and education in the SHS with internet non-use in the SHS.

The SHS 2019 dataset contains a table of 10,577 responses from each household surveyed. Within each household a random person was selected to answer individual level questions.

Of these questions we used the responses to the question: "Do you use the internet at all these days, either for your work or for your personal use?". There were 4,491 responses. We use this question to measure internet non-use.

Other response fields were used to measure age, income and education.

These are:

1. the individual respondent's age
2. the individual respondent's highest level of educational qualification.
3. Household-level response for net household income.

In order to correctly group responses from similar areas for analysis, we also used:

4. The Scottish Index of Multiple Deprivation[13] (SIMD) decile the household belongs to.
5. the local authority the household belongs to.

Only records where all fields have a response were selected for the dataset. Responses to level of education for "Other" and "Not known" were also removed, bringing the total number of records used for analysis down to 2,479.

In order to model area-level statistics the 2,479 individual level responses were grouped by local authority and then by SIMD decile. This grouping procedure was carried out to group responses from households from similar areas.

This produced 298 group-level data points where summary statistics were calculated based on the members of the group. Some of these groups contained fewer than 12 individual responses and so were added to neighbouring deciles within a local authority until all groups contained 12 or more responses. This reduced the number of groups to 135.

This dataset was then used to analyse the correlations between demographic factors and self-reported internet use.

[13]Scottish Index of Multiple Deprivation https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/

8.1 Coding Responses

Individual responses for internet use are coded to simplify answers into a binary variable. Table 2 below shows how responses to internet use were recoded. Those who use the internet at work only are included as internet users as they have the skills necessary to use the internet and access to an internet connection at work.

Table 2: Scottish Household Survey answers to Internet Use question

SHS Question	Response	Recode
net1	Both work and personal use	User
net1	No, never use the internet	Non-user
net1	Personal use only	User
net1	Work only	User

8.2 Categorising predictors

The summary statistics from the SHS dataset (age, income, and educational qualification) were chosen to match available Data Zone level variables. Having chosen age, income and education as potential predictors of digital exclusion, we considered what specific aspects of these variables we should use in further analysis. For example, we could use proportion of people over 55 years old, over 75 years old, or another proportion, to measure the effect of age.

Since Data Zones are geographical areas with varying numbers of people in them, using the raw values from the data sets would not produce usable results. Proportions of the population are used instead.

In order to produce population-level measures, we used the proportion of a population within a category as the measure for analysis. Age was split into cumulative brackets, detailed in Table 3. Then the number of people in a given Data Zone who are aged over 35, 55 and 75 were counted and then divided by the total number of people in the Data Zone aged 16 and over, producing the proportion of the adult population over a certain age. Income and Education were divided into sequential rather than cumulative brackets.

Education was measured using the Scottish Credit and Qualification Framework (SCQF) system, which allows categorisation and comparison of various equivalent Scottish qualifications[14].The full descriptions of the different qualifications included in the SCQF is given in the glossary at the end of this paper.

The variable categories used for digital exclusion analysis are shown in Table 3.

[14]Scottish Credit and Qualification Framework https://scqf.org.uk/about-the-framework/

Table 3: Categories of Digital Exclusion Measures

Variable	Category groups	Name for analysis
Age (years)	Age over 15 (16+)	AgeA
	Age over 35 (36+)	AgeB
	Age over 55 (56+)	AgeC
	Age over 75 (76+)	AgeD
Income (£)	0 – £12.5k	income1
	>£12.5k – 20k	income2
	>£20k – 30k	income3
	>£30k – 40k	income4
	>£40k +	income5
Education	No qualification	SCQF 0
	Standard Grade	SCQF 1
	Higher Grade	SCQF 2
	Higher National Diploma	SCQF 3
	University degree or higher	SCQF 4

Table 4: Variables used for digital exclusion correlation analysis

Summary Statistic	Categories	Description
Non-Users	1	Proportion of non-users in group
Age	3	Proportion of people in group aged over 35, over 55 and over 75
Income	4	Proportion of people in group with income less than £12.5k, £12.5-20k, >£20k-30k, >£30k-40k and equal to or over £40k
Education (using Scottish Qualification and Credit Framework levels)	5	Proportion of people in group with no qualifications, and with qualifications at different levels of education.

9. Estimating Digital Exclusion: Combination of demographic predictors

Beta regression was used to measure which variable categories were most closely correlated with non-use of the internet.

Several approaches including generalised linear modelling and principal component analysis were considered and explored to analyse the potential power of underlying demographic factors to predict digital exclusion. Several alternative data sources were also considered before the SHS data was chosen. Beta regression was chosen on the advice of the External Methodology Assurance Panel. Beta regression is a type of regression analysis that can be used where the data are expressed as a proportion; for example, the proportion of the population over 75, or the proportion of the population who do not use the internet.

The dataset on internet use was reformatted to use the same format as Data Zone level data for Scotland, in order to make the results applicable to the national-level datasets we wish to use to predict digital exclusion.

The SHS Data Zone data set (shown in table 3) was reduced after analysis to nine variables (shown in table 5). Initially, all the individual variables listed in table 3 and their interactions were used to create the model. The complexity of the model was then iteratively reduced by removing variables and interaction terms until only a set of significant predictors are left. The variables used in the finalised model are given below in tables 5 and 6.

The results of the beta regression showed that several variables had a strong correlation with internet non-use.

Table 5: Correlation and significance of variables used in the digital exclusion index model with internet non-use in the SHS dataset.

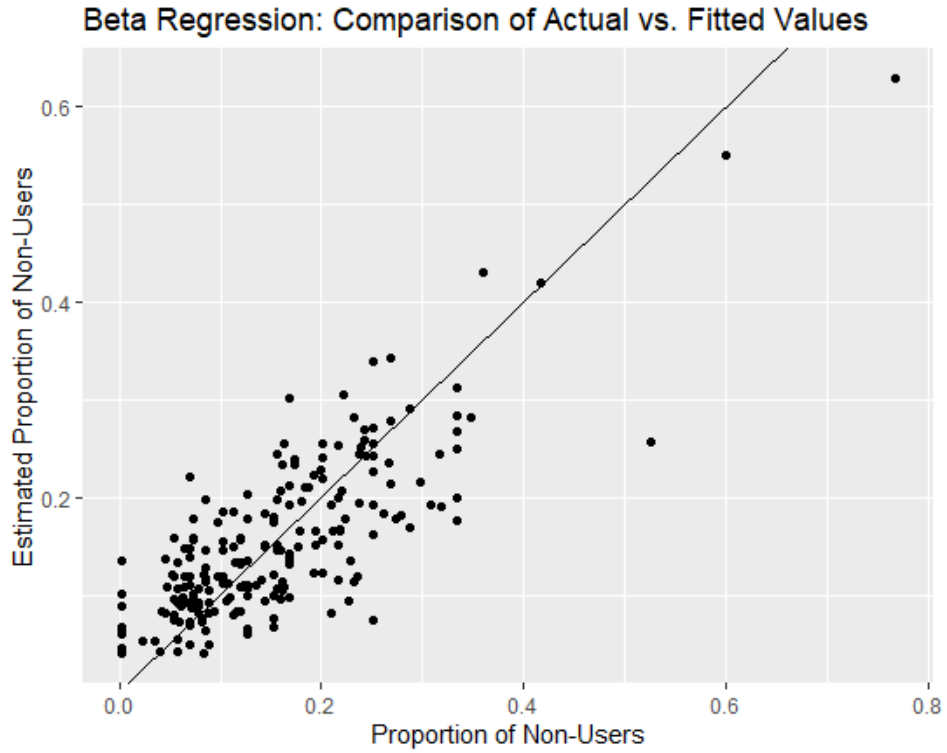
	Estimate	Standard Error	Z value	Pr(>|z|)
AgeD (over 75)	2.1658	0.5249	4.126	<0.001
AgeC (over 55)	1.8094	0.3474	5.209	<0.001
AgeB (over 35)	-0.813	0.393	-2.069	0.039
income1(0 – £12.5k)	0.8644	0.4178	2.069	0.039
income3 (>£20k-30k)	1.4451	0.4542	3.182	0.015
income5 (over £40k)	2.3274	0.5201	4.475	<0.001
SCQF 0	1.5729	0.4305	3.654	<0.001
SCQF 2	-2.24	0.5179	-4.325	<0.001
SCQF 4	-1.7628	0.3656	-4.822	<0.001

Table 6: Correlation and significance of variable interactions used in the digital exclusion index model with internet non-use in the SHS dataset.

	Estimate	Standard Error	Z value	Pr(>|z|)
Interaction between AgeD and income1 (AgeD)	2.776	1.281	2.167	<0.05
Interaction between AgeD and income1 (income1)	2.6054	1.0308	2.527	<0.05

Figure 1 shows the ability of the Digital exclusion model to estimate the proportion of internet non-users in an area, compared to the actual proportion of internet non-users reported in an area in the Scottish Household Survey.

Figure 1: Correlation of proportion of internet non users in SHS dataset compared with proportion of internet non users predicted by model



Objective ID :

10. Creating a national model of Digital Exclusion

We obtained national level datasets with information about the age, income and education in different Data Zones, as described in Table 1. We then transformed the Data Zone level data to planning area level, as planning areas are the geography used to plan the census. The digital exclusion index has been updated when new datasets have become available, and may be updated again if more up to date datasets are published prior to census and updating does not disrupt other processes.

10.1.1 Age

Using the Mid-2018 Small Area Population Estimates (published 2019) for the 2011 Data Zones data set, the estimated population by single year of age by 2011 Data Zone area was produced. This breaks down the estimated number of people living in each Data Zone by their age.

10.1.2 Qualifications

Data on qualifications has been taken from the Scotland's Census 2011 bulk output tables. This data set details the number of people aged 16 or over by the level of their highest educational or vocational qualification. This is split into 5 categories which closely match SCQF qualification levels. Adjustments were made to match census outputs to SCQF levels for modelling purposes. Further details can be found in the glossary.

10.1.3 Income

Local Level Distribution of Household Income Estimates 2014 gives the proportion of households in a Data Zone with weekly income under set intervals.

10.2 Using variables to produce digital exclusion measure

The age, income and education variables selected were then used to produce a predicted proportion of internet non-users for each planning area. A Beta Regression was performed with a logit link function to estimate the output variable. Since the method relies on a transformation of bounded variables into unbounded space, the records which have values of 0 or 1 were scaled into the interval. Values that are exactly 0 or 1 cannot be handled by the model, so these values were altered slightly in order to allow the use of this method.

This results in a model that takes Data Zone variables and predicts digital exclusion levels in planning areas from those variables. Planning areas were then ranked by their predicted level of digital exclusion to produce the Digital Exclusion Index (DEI).

11. Evidence for Digital Exclusion Index use from Census Rehearsal

11.1 Rehearsal data

National Records of Scotland carried out a major public rehearsal exercise that tested some of the systems and services that will be used in the 2022 Census.

The rehearsal took place between 7 October and 7 November 2019, using a reference day of 13 October, and was conducted in three local authority areas; namely parts of Glasgow City, Dumfries and Galloway and Na h-Eileanan Siar. These areas were selected to allow National Records of Scotland to test approaches in rural, urban and diverse communities. Participation in the rehearsal was on a voluntary basis. Over 72,000 households were contacted with nearly 18,500 responding. More detail on the rehearsal is available at https://www.scotlandscensus.gov.uk/news/census-rehearsal-evaluation-report-published

During rehearsal, data was collected on the paper request rates (the proportion of contacted households that requested paper questionnaires) and the paper return rates (the proportion of returns that were on paper questionnaires rather than on the default online questionnaire). These measures within each planning area were used in conjunction with planning area DEI scores to investigate and validate the DEI. The request of paper questionnaires is used as an indicator that a household is engaged with the census rehearsal process but is unwilling or unable to complete the online questionnaire.

Completing the rehearsal survey was voluntary, while filling out the census is not. Therefore rehearsal data does not predict return rates in the actual census, as voluntary respondents will be a smaller and demographically skewed sample of census respondents. We are assuming that patterns of response in rehearsal will broadly mirror those in Census as this has been the case in previous rehearsals and Censuses, even though request and return rates are very different.

11.2 Digital Access Map

The original DEI contained an access component, to reflect that a reasonable speed of internet access is needed for digital participation.

Comparing OFCOM broadband access data from 2016 with data from 2019 showed that there were significant changes at planning area level, with 402 planning areas showing an increase of 10 percent or more in the proportion of households receiving 10 Mbps. However, the access component of the DEI did not improve the predictive ability of the index. The reason for broadband access not predicting observed levels of paper return use could be due to:

- Mobile internet services may be used in areas with poor wired/cable broadband performance, allowing for increased internet access above what local broadband speed would indicate.

- The primary driver of digital exclusion could be whether a user is skilled/confident enough to use the internet or wants to use the internet, rather than whether fast broadband is available in their area.

We removed the broadband access component from the DEI entirely, due to broadband access adding no value to the prediction of digital exclusion markers in rehearsal data. Alternative data sources for internet access through mobile networks has been sought but is not available. The proposed DEI will be comprised of only the beta regression model for estimating the proportion of non-users in an area.

12. Improving segmentation of the Digital Exclusion Index

In this section different partitions are considered for their ability to produce five groups of planning areas which have increasing paper return rates (the proportion of returns that were on paper questionnaires rather than on the default online return questionnaire).

The partitions are intended to group planning areas into five groups. Group 1 should contain the least digitally excluded part of the population and group 5 the most digitally excluded part of the population. These groups should reflect increasing expression of markers of digital exclusion. The three different partitions considered are given in Table 7.

Table 7: Three potential partitions for grouping digital exclusion scores

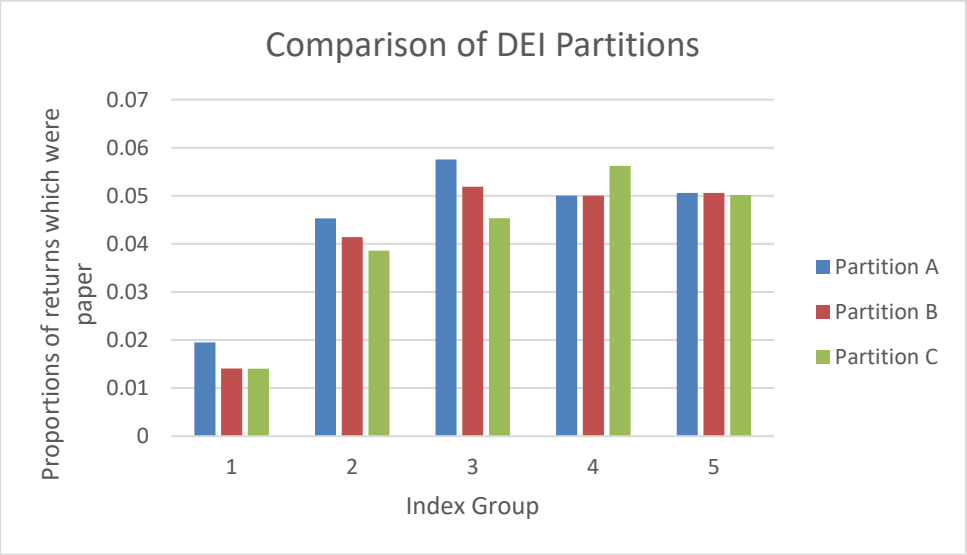
Partition	1	2	3	4	5
A	40%	40%	10%	8%	2%
B	30%	30%	30%	8%	2%
C	25%	25%	25%	15%	10%

Rehearsal results for paper return rates and paper request rates were used as a proxy for digital exclusion, as households which are digitally excluded are more likely to request and return paper questionnaires. This is not an ideal indicator as areas with lower engagement with census that are also digitally excluded may have low paper requests rates due to lack of engagement with census rather than higher digital skills.
Especially in the voluntary census rehearsal, digitally excluded people who also have low engagement with census may not respond at all rather than take the time to request a paper questionnaire, complete and return it. Therefore paper returns and requests are used as broad indicators of relative digital exclusion only.

All planning areas were assigned index values according to the three partition sets outlined above. Then the rehearsal planning areas were grouped by their index values and the paper return rates are calculated for each group, as the proportion of returns received on paper questionnaires rather than online.

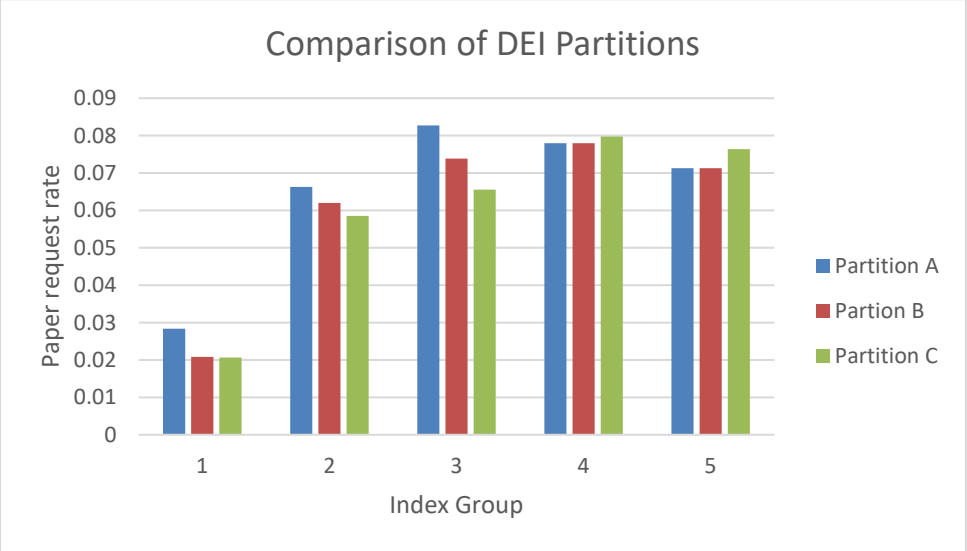
The results are visualised in Figure 2 below.

Figure 2: Bar Graph of Proportion of Paper Rehearsal Returns grouped by Proposed Partitions



In addition to the paper return rate, the paper request rate was also used as an indicator of digital exclusion. The paper request rate is shown in Figure 3 below. The calculation for paper request rate is the number of paper requests divided by the number of eligible households in the planning area. This presents a potential issue where a household may have submitted multiple requests for paper questionnaires, which would produce a slight over-count of the actual number of households requiring paper. Because the number of cases is assumed to be small the impact is assumed to be negligible.

Figure 3: Bar Graph of Paper Request rates grouped by Proposed Partitions



Objective ID :

12.1 Choice of partition option

Option C best reflected a steadily increasing level of paper return and request rates with increasing digital exclusion index group and was chosen for the digital exclusion index. The segmentation is shown in Table 10. The groups of this proposed partition have the closest pattern to a steadily increasing proportion of returns by paper, except for group 5. However the difference between group 4 and 5 is small enough that it could be coincidental.

Table 10: Segmentation of the DEI

Name	Percentage of Planning Areas	Description
DEI 1	25%	Least digitally excluded
DEI 2	25%	
DEI 3	25%	
DEI 4	15%	
DEI 5	10%	Most digitally excluded

13. Strengths and limitations

13.1 Strengths

The DEI is based on demographic variables (age, education and income) that are meaningful at a geographic area level and that do not change rapidly within areas over time. The demographic variables chosen also link to identified barriers to digital participation (interest, skill and cost).

The DEI has been tested in the field in NRS census rehearsal 2019 and improved as a result of testing in order to better reflect observed digital exclusion markers. The methodology behind this index has been reviewed by an external academic panel. An older dataset was removed from analysis and a different methodology was used to select underlying variables and construct the index as a result of their recommendations.

13.2 Limitations

Digital exclusion is a rapidly evolving field of study, and levels of digital access and digital skills in the UK population have changed rapidly over recent years. The DEI and the data used in the above analysis was produced before the coronavirus pandemic, which has further altered the landscape of digital exclusion. However, the data needed to fully understand changes in digital exclusion will not be available in time to do confirmatory analysis prior to census. The DEI in its current form is likely to still have high predictive power, as the demographic factors on which it is based still predict likely barriers to digital participation (interest, skill and cost).

The DEI is produced at area level, rather than individual level, so it does not allow resources to be targeted to individual digitally excluded people within areas. It does

allow areas to be ranked by how digitally excluded the collective population within that area is expected to be. Individual-level targeting would be suitable for some potential digital exclusion interventions but not others, and would require obtaining current, complete, individual-level data across the whole of Scotland and conducting extensive further analysis.

The inclusion of pandemic-era (2020) digital exclusion data and other additional confirmatory analysis suggested by the external methodology review panel was not carried out, as much of the data necessary for such analysis was not collected or made available within the timescale for production of this index.

14. Conclusion

The Digital Exclusion Index (DEI) will provide guidance on which census planning areas contain the highest proportion of digitally excluded people. The index will allow geographically targeted interventions to mitigate digital exclusion to be implemented in areas where they will reach higher proportions of digitally excluded people.

The DEI is based on demographic characteristics that are linked to digital exclusion and that are recorded in good quality datasets that cover the whole of Scotland at a small geographic area level.

15. Glossary and acronyms

Acronym	Term	Definition
DEI	Digital Exclusion Index	An index estimating the relative degree of digital exclusion (lack of Internet access and/or persons lacking digital skills) within a Planning Area. The Digital Exclusion Index is created from a ranked list of planning areas in order of the estimated proportion of digitally excluded people.
DZ	Data Zone	Data Zones are the core geography for dissemination of results from Scottish Neighbourhood Statistics (SNS). The Data Zone geography covers the whole of Scotland and nest within local authority boundaries (as they were in 2011). Data zones are groups of 2011 Census output areas which have populations of around 500 to 1,000 residents.
HtC index	Hard to Count index	An index indicating how willing households within a Planning Area will be to respond to the census. The Hard to Count Index is created from a ranked list of planning areas based on factors found to be associated with Census non-response.

NISRA	Northern Ireland Statistical Research Authority	NISRA is the principal source of official statistics and social research on Northern Ireland. They conduct the census in Northern Ireland. https://www.nisra.gov.uk/
OFCOM	Office of Communications	Ofcom is the regulatory body for UK communications services. https://www.ofcom.org.uk/about-ofcom
ONS	Office for National Statistics	The UK's largest independent producer of official statistics and its recognised national statistical institute. It is responsible for collecting and publishing statistics related to the economy, population and society at national, regional and local levels. ONS also conduct the census in England and Wales every 10 years. https://www.ons.gov.uk/aboutus
PA	Planning Area	Planning Areas (PAs) are geographic areas built from groups of postcodes and averaging between 200-400 residential addresses, although there are a small number outside that range. PAs nest within Local Authorities and will be used during the production of the Hard to Count Index and for the prioritisation of enumeration addresses to follow-up.
SCQF	Scottish Credit and Qualification Framework	Level of education as recorded in the census is measured with the Scottish Credit and Qualification Framework system. In Scotland's Census 2011 Bulk outputs: Education Level, the highest level of qualification is defined as in the SCQF: • Level 0: No qualifications. • Level 1: O Grade, Standard Grade, Access 3 Cluster, Intermediate 1 or 2, GCSE, CSE, Senior Certification or equivalent; GSVQ Foundation or Intermediate, SVQ level 1 or 2, SCOTVEC Module, City and Guilds Craft or equivalent; Other school qualifications not already mentioned (including foreign qualifications). • Level 2: SCE Higher Grade, Higher, Advanced Higher, CSYS, A Level, AS Level, Advanced Senior Certificate or equivalent; GSVQ Advanced, SVQ level 3, ONC, OND, SCOTVEC National Diploma, City and Guilds Advanced Craft or equivalent.

		• Level 3: HNC, HND, SVQ level 4 or equivalent; Other post-school but pre-Higher Education qualifications not already mentioned (including foreign qualifications) Level 4 and above: Degree, Postgraduate qualifications, Masters, PhD, SVQ level 5 or equivalent; Professional qualifications (for example, teaching, nursing, accountancy); Other Higher Education qualifications not already mentioned (including foreign qualifications).
SHS	Scottish Household Survey	The Scottish Household Survey (SHS) is an annual survey of over 10,000 households. The survey covers a wide range of topics relating to the composition, characteristics, behaviour and attitudes of Scottish households and adults http://scottishhouseholdsurvey.com/
SIMD	Scottish Index of Multiple Deprivations	The Scottish Index of Multiple Deprivation is a relative measure of deprivation across 6,976 small areas (called Data Zones). If an area is identified as 'deprived', this can relate to people having a low income but it can also mean fewer resources or opportunities. SIMD looks at the extent to which an area is deprived across seven domains: income, employment, education, health, access to services, crime and housing. https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/
SVQ	Scottish Vocational Qualification	The SVQ is a work-based qualification which recognises the skills, knowledge and understanding a person needs to do a job.