

Article

Digital Propensity Index for England and Wales LSOAs: Census 2021

How Census 2021 online response rates, combined with modelled predictions, created a comparative measure of digital uptake for England and Wales at Lower layer Super Output Area (LSOA) level.

Contact:
Census customer services
census.customerservices@ons.
gov.uk
+44 1392 444972

Release date:
8 February 2023

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [About the online-first census](#)
3. [About the Digital Propensity Index](#)
4. [Predicted paper-first online shares](#)
5. [Digital Propensity Index data](#)
6. [Glossary](#)
7. [Data sources and quality](#)
8. [Future developments](#)
9. [Related links](#)
10. [Cite this article](#)

1 . Main points

- For Census 2021, households were sent either an initial contact letter containing an online access code ("online-first"), or a paper questionnaire ("paper-first"); we determined the Lower layer Super Output Areas (LSOAs) that would be treated as paper-first using two hard-to-count indices, Hard-to-Count Willingness and Hard-to-Count Digital.
- The Digital Propensity Index (DPI) is a unique, rich, and relevant data source of digital uptake for LSOAs across England and Wales.
- The DPI combines, with associated measures of uncertainty, the actual online share of responses observed for online-first LSOAs and the predicted online share of responses for paper-first LSOAs, if they had been online-first.
- Of the households sent a paper questionnaire (including an access code to respond online) in Census 2021, 46.4% responded online; the equivalent proportion was 92.5% when we modelled predictions using the DPI.
- The DPI, a measure of how confident households are using government online resources, is an alternative to the Hard-to-Count Digital Index which was used in planning Census 2021.
- These results potentially imply that future research could more confidently take an online approach to data collection, although we recognize that not everyone will be capable of responding, or willing to respond, online.

There is always uncertainty and error in a model's predictions. Read how we have accounted for this in "Interpreting the data" in [Section 7: Data sources and quality](#).

2 . About the online-first census

Census 2021 was the first online-first census in England and Wales where all households received a unique access code (UAC) to complete the census online. Most households only received a UAC and are referred to as online-first areas. However, to maximise response, 11% of households (50% of Welsh households and 9% of English households) first received a paper questionnaire and a UAC. These are referred to as paper-first areas.

We identified Lower layer Super Output Areas (LSOAs) rather than individual addresses as needing to be paper-first. This is because we recognise there will be some similarity between an area's households and residents. Additionally, designing which areas would be paper-first was based on data at 2011 LSOA level. Specifically, these were areas where the take-up of the online option was expected to be low, but willingness to take part without further prompts (reminder letters or field visits) was high.

Online share

Census 2021 had quality targets specifically linked to it being the first online-first census. We exceeded a target of achieving at least 75% of responses to Census 2021 online. In total, 88.9% of responding households in England and Wales chose to respond online.

Specifically, paper-first [LSOAs](#) had an online share of 46.4%, whereas online-first LSOAs had an online share of 94.2%.

However, the Census 2021 collection design influenced the online share. The online share is the proportion of responding households that have responded online. It is not the proportion of all households in England and Wales (including non-responding households) that have responded online.

Neither are these figures about overall response, nor does a high digital uptake in an area necessarily correspond to a high overall response rate. For example, urban areas often had a lower overall response for Census 2021, but a higher proportion of the responding households in these areas completed an online census form.

Read more about the share of households responding online, and how the collection design influenced it, in [our Designing a digital-first census article](#).

3 . About the Digital Propensity Index

We created the 2021 Digital Propensity Index (DPI) to give a relative measure of digital propensity at Lower layer Super Output Area (LSOA) level across England and Wales. This provides additional value from Census 2021 data by providing a valuable tool for anyone planning digital services or researching digital deprivation.

To do this, we predicted the online share for paper-first areas had they been online first. In doing so, we provided a comparative measure of digital propensity at LSOA level across England and Wales, independent of the census paper strategy. Combining these modelled predictions with the actual online shares for online-first areas gives us DPI scores for every LSOA (as defined following the 2011 Census) and local authority in England and Wales.

You can now download the DPI along with associated measures of uncertainty at LSOA level.

4 . Predicted paper-first online shares

The paper-first Lower layer Super Output Areas (LSOAs) were all predicted to have had an increased online share, if they had been online-first areas. See Section 7: Data sources and quality for how the predictions were made. This is expected as our strategy categorized whole LSOAs as paper-first rather than individual addresses. However, as Census 2021 shows more generally there is variation across residents even in small areas. The increased online share of paper-first areas saw their average online share increase by more than 40 percentage points, as shown in Table 1. For comparison, the mean online share of online-first areas was 94.2%.

Table 1: The mean and median breakdown for the paper-first area predictions, all England and Wales LSOAs, Census 2021

	Predicted online share	Actual online share	Difference in percentage points
Mean	92.5%	46.7%	45.8%
Median	92.5%	45.3%	47.2%

Source: Office for National Statistics

Notes

1. LSOAs are Lower layer Super Output Areas.

We calculated the online share at local authority (LA) level combining the achieved online share for online-first LSOAs and the predicted online share for paper-first areas. We did this by calculating a weighted average taking into account the number of households in each LSOA, rather than simply averaging the LSOA predictions in each LA.

Table 2 shows the 10 LAs with the highest predicted online share.

Table 2: The 10 local authorities with the highest Digital Propensity Index, aggregated from LSOA level, Census 2021

Local authority name	Region	Percentage of LA that is paper first	Actual online share	Digital Propensity Index (predicted online share if all areas online-first)
Tower Hamlets	London	3.8%	96.4%	97.3%
Newham	London	1.0%	96.7%	97.0%
City of London	London	0.0%	96.7%	96.7%
Hounslow	London	0.0%	96.6%	96.6%
Slough	South East	6.8%	94.6%	96.6%
Wandsworth	London	3.4%	95.3%	96.5%
Bracknell Forest	South East	2.1%	95.6%	96.4%
Lambeth	London	0.0%	96.3%	96.3%
Southwark	London	0.6%	96.1%	96.3%
Harrow	London	0.0%	96.3%	96.3%

Source: Office for National Statistics

Notes

1. LSOA is Lower layer Super Output Area.
2. Percentages are rounded to 1 decimal place.

Table 3 shows the 10 LAs with the lowest predicted online share.

Table 3: The 10 local authorities with the lowest Digital Propensity Index, aggregated from LSOA level, Census 2021

Local authority name	Region or country	Percentage of LA that is paper first	Actual online share	Digital Propensity Index (predicted online share if all areas online-first)
North Norfolk	East of England	17.4%	80.8%	90.6%
East Devon	South West	24.0%	78.0%	91.0%
Tendring	East of England	20.0%	80.9%	91.0%
Isle of Wight	South East	21.0%	80.3%	91.1%
Rother	South East	7.8%	87.0%	91.1%
East Lindsey	East Midlands	34.0%	75.9%	91.2%
Great Yarmouth	East of England	13.7%	84.3%	91.2%
Powys	Wales	39.5%	69.7%	91.3%
Torbay	South West	8.7%	86.9%	91.3%
Scarborough	Yorkshire and The Humber	10.8%	86.0%	91.5%

Source: Office for National Statistics

Notes

1. LSOA is Lower layer Super Output Area.

Figure 1: Digital Propensity Index: predicted Census 2021 online share of responses, if all LSOAs were online first, by 2020 local authority

Notes:

1. LSOAs are Lower layer Super Output Areas.

Download the data

[.xlsx](#)

To maximise response, as described in [the UK Statistics Authority's Maximising Response Strategy Overview \(docx, 289KB\)](#), in Wales, 50% of households were paper first. In comparison, in England, only 9% of households were paper first. So, the paper-first increases have a greater impact on Wales and Welsh local authorities. This can be seen in Table 4 as 8 of the 10 local authorities (LAs) with the biggest online share increase are from Wales.

Table 4: The 10 biggest local authority online share increases, aggregated from LSOA level, Census 2021

Local authority name	Region or country	Percentage of LA that is paper first	Actual online share	Digital Propensity Index (predicted online share if all areas online-first)	Difference in percentage points
Gwynedd	Wales	72.8%	53.8%	92.2%	38.4%
Blaenau Gwent	Wales	70.6%	55.0%	91.9%	36.9%
Denbighshire	Wales	63.0%	58.2%	91.7%	33.5%
Wrexham	Wales	64.9%	60.3%	93.0%	32.7%
Isle of Anglesey	Wales	61.4%	59.8%	92.1%	32.3%
Conwy	Wales	59.0%	59.7%	91.6%	31.9%
Boston	East Midlands	60.4%	61.0%	92.2%	31.2%
Rhondda Cynon Taf	Wales	63.5%	61.7%	92.6%	30.9%
Torfaen	Wales	61.7%	62.5%	92.7%	30.2%
Kingston upon Hull, City of	Yorkshire and The Humber	58.9%	63.8%	92.7%	28.9%

Source: Office for National Statistics

Notes

1. LA is local authority.
2. LSOA is Lower layer Super Output Area.

A total of 50 LAs have no notable difference (less than 0.1%) between their actual and predicted online share. Less than 0.1% of households in these LAs were paper first so the paper predictions did not have a big influence.

The LA results have been produced using 2020 LAs to allow for direct comparison with previously published research in our [Designing a digital-first census article](#) on the achieved online share for Census 2021. Figure 1 is based on the LAs in 2020 and the results are available to download.

The LA results have also been produced using the 2021 [restructuring of local government](#). Figure 2 is based upon the LAs in 2021 and the results are available to download.

Figure 2: Digital Propensity Index: predicted Census 2021 online share of responses, if all LSOAs were online first, by 2021 local authority

Download the data

[.xlsx](#)

Table 5: Actual and predicted online shares, Wales and English regions, Census 2021

Region or country	Actual online share	Digital Propensity Index (predicted online share if all areas online-first)	Difference in percentage points
East Midlands	84.7%	93.7%	9.0%
East of England	91.3%	94.1%	2.8%
London	95.5%	95.9%	0.4%
North East	91.0%	93.3%	2.3%
North West	92.1%	93.9%	1.8%
South East	93.2%	94.5%	1.3%
South West	86.3%	93.0%	6.7%
Wales	67.6%	92.6%	25.0%
West Midlands	83.6%	93.4%	9.8%
Yorkshire and The Humber	88.0%	93.4%	5.4%

Source: Office for National Statistics

Notes

1. LSOAs are Lower layer Super Output Areas.

The Digital Propensity Index for Wales and English regions is shown in Table 5. Wales has seen much larger increases than other regions because of the high proportion of paper-first LSOAs in Wales. Like the LA estimates, the calculation uses a weighted average considering the number of households in each LSOA, rather than simply averaging the LSOA predictions in each region.

Predictions show greater digital propensity

The mean online share for online-first areas is 94.2%. The predicted mean online share for paper-first areas is 92.5%. Across England and Wales the predicted online share is 94%.

The paper-first areas were designed as such because they were expected to have a lower digital uptake. The predictions from the model indicate that many of the paper-first areas would still have achieved a lower online share than online-first areas. Wales, with a higher proportion of paper-first areas than any English region, was still predicted to have the lowest overall online share, despite the model predicting the largest increase in digital uptake. Also, the predicted mean for paper-first areas is 92.5%, whereas the mean for online first areas is 94.2%.

These predictions imply that many of the paper-first areas do have a relatively low digital uptake. This is despite the predictions resulting in the average online share of paper-first areas increasing by more than 40 percentage points.

Therefore, the modelling broadly aligns with the information used in the collection design of Census 2021 because the paper-first areas have a relatively lower predicted online share. However, the predicted mean online share for paper-first areas of 92.5% is still much higher than the actual online share received in said areas. This suggests that future data collection could take a greater online-first approach.

Please note, there is always uncertainty and error in a model's predictions. We recommend using confidence intervals provided with the LSOA predictions to decide whether one LSOA has a notably higher digital propensity than another. You can find more information on managing the model's uncertainty in "Interpreting the data" in [Section 7: Data sources and quality](#).

5 . Digital Propensity Index data

[Digital Propensity Index for Census 2021 at Lower layer Super Output Areas \(LSOAs\), England and Wales](#)
Dataset | Released 8 February 2023

Digital Propensity Index scores and associated confidence intervals for LSOAs as defined in 2011 in England and Wales.

[Digital Propensity Index for Census 2021 at local authority, region and country level, England and Wales](#)
Dataset | Released 8 February 2023

Digital Propensity Index scores for local authorities, as defined in December 2020 and December 2021, regions and countries in England and Wales.

6 . Glossary

Digital Propensity Index

A measure of how often individuals use communication technology.

Hard-to-Count Index

A measure of the relative willingness of residents in an area to respond to the census without further prompts and the relative likelihood that they will respond online. The Hard-to-Count Index is made up of the Hard-to-Count Digital Index and the Hard-to-Count Willingness Index, both of which are used as variables in the model.

Lower layer Super Output Area (LSOA)

Lower layer Super Output Areas (LSOAs) are made up of groups of Output Areas (OAs), usually four or five. They comprise between 400 and 1,200 households and have a usually resident population of between 1,000 and 3,000 persons. They were first created following the 2001 Census and may change after each census.

Online-first area

LSOAs where households first received only a letter with an access code to complete the census online.

Paper-first area

LSOAs where households first received a paper questionnaire and an access code to complete the census online.

Binomial logistic regression

A logistic regression model for a binomial dependent variable with y successes out of n trials and one or more independent variables.

7 . Data sources and quality

Measuring the data

We used a binomial logistic regression for the Lower layer Super Output Area (LSOA) predictions. This is because it allowed us to model how multiple independent variables affect the likelihood that households will respond online. We created two models at LSOA level, an English and a Welsh model, to account for the respective indices of multiple deprivation from each country.

We also used the Census 2021 online share for online-first areas as the dependent variable because we aimed to predict the online share of returns. The dependent variable is not binary in the sense that it is not on a 0/1 scale. However, the dependent variable was binomial in that y is defined as the number of households in each LSOA responding online and n is defined as the total number of responding households in each LSOA.

The independent variables included in the models are:

- age - the proportion of household reference persons in the age group 65 years and over at LSOA level from the 2011 Census data; the variable was logit transformed to better meet model assumptions, and 2011 Census data were used as at the time of modelling age was not available at LSOA from Census 2021
- Hard-to-Count Digital Index (HtC D) - an index from 1 to 5 (5 being the hardest to count), showing the relative propensity of households in an area to respond to Census 2021 online (HTC D 1 is the reference category)
- Hard-to-Count Willingness Index (HtC W) - an index from 1 to 5 (5 being the hardest to count), showing the relative willingness of households in an area to respond to Census 2021 within 10 days after Census Day (HTC W 1 is the reference category)
- urban/rural classification - urban/rural as defined by the [Official Statistics 2011 Rural Urban Classification](#), which states whether the LSOA is classed as urban or rural (rural is the reference category)
- region - region as defined by Eurostat's [Nomenclature of Territorial Units for Statistics \(NUTS\) in the UK](#) (London is the reference category), included in the English model
- English Index of Multiple Deprivation (IMD) - included only in the English model, an index from 1 to 10, showing the relative deprivation for an LSOA (IMD 1 is the reference category)
- Welsh Index of Multiple Deprivation (WIMD) - included only in the Welsh model, an index from 1 to 10, showing the relative deprivation for an LSOA (IMD 1 is the reference category)

We created and trained both models using the data from the online-first areas. Then, we applied the models to the respective paper-first areas to produce the predicted online share for paper-first areas, had they been online first.

We used all the Welsh online-first areas as the sample for the Welsh model. The sample size for the Welsh model was 934 LSOAs.

We used the variable region in the English model. To make the sample more representative of what we were predicting, we made the sample match the regional breakdown of the paper-first areas. For example, 26.3% of paper-first areas in England are from the West Midlands, so 26.3% of the online sample we used in the model were from the West Midlands.

To achieve this regional representation, the sample size for the English model was 9,998 LSOAs of the 29,835 online-first LSOAs in England. Sensitivity analyses were conducted to ensure that the sample of LSOAs selected for the English model had minimal impact on the final predictions.

We carried out checks to ensure the assumptions underpinning the models were met. The checks we used were:

- the Cook's distance to ensure no LSOAs were having an undue impact on the final predictions
- Variance Inflation Factors (VIFs) were assessed to ensure there was no severe multicollinearity
- linearity of continuous independent variables on the logit scale was assessed
- residual plots

Then, we used the online share from online-first areas and the predicted online share from paper-first areas to create the Digital Propensity Index (DPI).

The LSOA online shares were also aggregated to find local authority and regional results for quality assurance and publishing. This is done by calculating a weighted average taking into account the number of households in each LSOA, rather than simply averaging the LSOA predictions in each LA.

Interpreting the data

When interpreting the results, it is important to remember that 11% of the LSOAs across England and Wales were paper first. As such, 11% of the LSOA results are predictions produced from the models. There is always some level of error and uncertainty in a model's predictions as the model is limited by the data used, and not all factors can be considered.

To reflect this, we have published confidence intervals (CIs) with the paper-first predictions and calculated coefficients of variation (CVs). These CIs and CVs show how much uncertainty there is in the modelled values, and the CI's give a range of values around the estimate which likely contains the correct value. Note that it is not possible to capture all forms of uncertainty in the confidence intervals. For example, the confidence intervals for the English models do not capture the uncertainty introduced from using a subset of LSOAs in the final model. Table 6 shows the CV results from each model. Often a CV under 20% is acceptable and the highest CV for both models is below 20%, with the mean well below 20%.

Table 6: The coefficient of variation (CV) results from the predictions of the Welsh model and English model, England and Wales LSOAs, Census 2021

	Welsh model	English model
Highest CV	6.7%	1.2%
Mean CV	2.7%	0.7%
Median CV	2.8%	0.7%

Source: Office for National Statistics

Notes

1. LSOAs are Lower layer Super Output Areas.

8 . Future developments

This data on the relative digital propensity of households at Lower layer Super Output Area (LSOA) level across England and Wales is useful for research and planning of digital services. These results potentially imply that future research could more confidently take an online approach to data collection, which is important with more of our surveys going online. However, we recognize that not everyone will be capable of responding, or willing to respond, online.

9 . Related links

[Designing a digital-first census](#)

Article | Released 4 October 2021

How we designed a digital-first census to make sure everyone could take part and the response to this.

10 . Cite this article

Office for National Statistics (ONS) released 8 February 2023, ONS website, article, [Digital Propensity Index for England and Wales LSOAs: Census 2021](#)