ONS Census 2021 consultation

Health Foundation response to new ONS consultation on Census 2021 Outputs

4 October 2021

Background

The Census is a survey of every household in England and Wales taken once every 10 years. The latest Census was taken during March 2021. Census data is used extensively by public and private sector organisations, providing information on the population on Census day, forming the basis for other statistics, and informing planning decisions for the future.

The Office for National Statistics are consulting on user needs for Census data to inform their data outputs. The ONS aim to increase the use of Census data through this consultation. They are seeking feedback on proposals and wish to understand user needs for Census data. This is particularly important around topic areas such as where and how we live, work and study, that have experienced an increased rate of change since March 2020.

Below is the current proposal for the output release schedule. ONS guidance suggests that providing data at a finer geographical detail will be prioritised before multivariate tables.

From March 2022 to March 2023 they plan to release most of the data, but it will be in three phases:

1) Initial findings, population estimates, and then topic summaries for single variables ie general health down to small areas such as output areas

2) Start to release cross-tabulated data ie general health by ethnic group in their new build your own tool, or ready-made tables

3) Release data about alternative population bases and more detailed analysis tables

From March 2023 they hope to release UK data and other more bespoke analysis such as new healthy life expectancy estimates for small areas.

Approach

Content

The Census 2021 data will help to inform our understanding of the population, the wider determinants of health, and shape views on health and health inequalities in England. Its development is therefore of importance to the analytical and strategic work of the Health Foundation.

For the Healthy Lives team, Census data will inform wider topic analysis in our publications and will feature on the evidence hub as standalone analysis pages. The data will also

support several other programmes, including: the cross-government inequalities strategy, the Social and Economic Value of Health (SEVH), place-based funding programmes (Shaping Places for Healthier Lives, Economies for Healthy Lives) and numerous projects focused on improving the public health system.

The questionnaire is structured into seven sections (outlined below), each of which have several subtopics and questions to respond to.

1. **Why are we consulting (no questions)**
2. **The shape of the outputs and analysis release schedule:**
    a. Geographical boundaries
    b. Topic-based summaries and area profiles
    c. Multivariate data
    d. Data user needs for UK data
3. **Main changes to variables compared to the 2011 Census:**
    a. Sexual orientation and gender identity
    b. Armed forces veterans and proxy response
    c. New response options for ethnic group
    d. Age and country of birth classifications
4. **Proposals for feasibility work to derive new variables:** proposed new variables or indicators, such as:
    a. Education and employment: student populations, NEET, key or critical worker, economic risk from pandemic (occupation based)
    b. Health and living arrangements: COVID-19 health risk indicator ie extremely clinically vulnerable, HMOs/multigenerational households
    c. Accommodation types and vacant addresses: care homes, homelessness
    d. Additional data needs
5. **Population-base specifications:**
    a. Alternative population bases (families, workplace, workday)
    b. Small population data (detailed ethnic and country of birth)
6. **Taking a census in a period of change:**
    a. Place of residence
    b. Economic activity, employment
    c. Travel to work and other topics
    d. Impacts on other outputs
    e. Emerging data needs
7. **Paradata:**
    a. Households and individual response rates
    b. Mode and language of completion

It seems sensible to only respond to the elements of the consultation which are most relevant to informing our work. In particular, we are interested in commenting on any questions which:

- refer to variables linked to the work within the organisation and in particular the wider determinants of health
- proposals on the new Census derived variables which should add value to our work and allow us to identify new trends and patterns
- discuss what the impact of taking a Census in a period of change will mean for future use of Census data, including implications for the devolved nations.

The consultation is of relevance to numerous teams within the organisation, and we have sought input from across the organisation during the process to ensure that our response is

well-aligned with the whole organisation's views. We have gathered perspectives from external partners and findings from these discussions are used to holistically inform the Health Foundation's consultation response.

Questions for response

Section 2: Shape of the outputs and analysis release

2a Geography

- All Census 2021 outputs, in particular ward boundaries, will reflect the geographies in place in May 2022, rather than as of Census Day, 21 March 2021

- ONS will be prioritising release of key admin and statistical small area geographies, such as output areas (OAs) and lower super output areas (LSOAs), before parliamentary constituencies and health areas. This will include merging and splitting of OA boundaries, and targeted realignment of OAs to ward 2022 boundaries.

- For Census 2021, the guidance will suggest using higher levels of geography to build other geographies where possible, rather than adding all individual OAs together. This is because the small cell key perturbation method that they will apply to Census 2021 data has less impact on larger areas.

Are your data needs met by the publication of statistics aligned with geographical boundaries in place as of May 2022?

Yes.

Are you content with the ONS's proposals to provide guidance on how to create best approximations of data for new geographical areas?

Yes, but we will need detailed guidance on how to best do this. ONS need to ensure that guidance is clear so that everyone uses the method consistently.

2b: Topic-based summaries and area profiles

The plan is to first release topic-based summaries in univariate tables at all geographies in the following topic order:
1. Demography and migration
2. Ethnic group, national identity, language and religion
3. Sexual orientation and gender identity
4. Health, disability and unpaid care
5. Housing
6. Labour market and travel to work
7. Education
8. Armed forces veterans

Do you agree with the recommended ordering of the topic-staggered approach to phase one?

We agree that population and demography topic should be published first as these will give us population bases that are critical to other work.

We are more likely to use the second topic (ethnic group, national identity, language and religion) to breakdown other variables such as health or employment, rather than on its own, but we do not disagree with its position in the list. This is because it will be very important for users understanding where communities are concentrated.

Next in the list is the sexual orientation and gender identity topic, the variables for which are yet to be classified (as shown in the 'Draft proposals for outputs content' Excel file). This could mean a longer processing time, therefore potentially slowing down the release of other topics. Although we are very interested in seeing the results for this new topic, ONS should ensure that it does not hinder other release times by being published ahead of them.

If this looks to be the case, we would like to see health, disability and unpaid care published as soon as possible after ethnicity, national identity, language and religion.

We consider information on housing, labour market, travel to work and education equally as the next highest priorities. These would provide greater understanding of variation in social and economic outcomes at a local level.

Overall, we broadly agree with the topic ordering, but want to emphasize that will be able to gain a much deeper understanding of the data through cross-tabulation of more than one variable. We will ultimately support the order that enables the process to progress most efficiently.

Do the table specifications for phase one meet your data user needs? Specifications here

Yes.

We would also be open to discussing any support we were able to provide that could enable more in-depth analysis of small area data, as this is something that is critical to our award holders.

Do the proposals for the content of area profiles meet your data user needs? Proposals here

Yes.

Do you require additional less detailed profiles for small parishes that will be merged with other areas in the standard profiles and tables?

No.

2c: Multivariate data

- The model for 2021 will be somewhat different than for 2011. ONS are introducing functionality for users to build their own tables, alongside the availability of a set of prescribed ready-made tables.

- Tables will not include totals or subtotals. We will apply the small cell key perturbation method to Census 2021 data as part of the disclosure control methodology

Do the table specifications for 'ready-made' tables meet your needs? Specifications here

Yes.

Do you plan to use the 'build-your-own' tables from flexible datasets functionality?

Yes, although, at the moment it is difficult to envisage how we can use it from the 'Draft proposals for outputs content' document. It would be very helpful for users to see a prototype before it is released. We would be very interested in understanding whether this tool will be replicated for 2011 data, or if there are plans to simplify comparative analysis between 2021 and 2011.

Do the dataset specifications to 'build-your-own' tables from flexible datasets meet your needs? Specifications here

Yes.

Do the plans for Census 2021 analysis meet your needs? (proposals linked below)

https://consultations.ons.gov.uk/external-affairs/census-2021-outputs-consultation/user_uploads/analysis_programme_proposals.pdf

Yes, we particularly welcome the release of 'HDUC-09: Spatial and socio-economic inequalities in health-state life expectancies', as these are long awaited and will build on the foundation of knowledge we have accumulated around the wider determinants of health. We ask that these are prioritised as part of the proposed publications beyond March 2023 or are brought forward if possible.

2d: Data user need for UK data

- ONS are responsible for collating data from across the four nations of the UK to produce UK Outputs. These outputs provide totals for the UK, as well as comparable data for small areas across the UK.

- England, Wales, and Northern Ireland had their Census in March 2021, but Scotland is having theirs in March 2022. This means that there will be a one-year difference in reference dates, and that the first results from Scotland will not be published before March 2023.

Do you plan to use UK level data or comparable data for small areas across the UK?

i.e. what data would you use and what for, variables and geography

We would use a variety of data at a UK level and for small areas across the UK such as: population, ethnicity, health, disability, unpaid care, employment, housing, and education. At a UK level, we would be able to understand trends in UK outcomes for these variables compared with 2011. We would use data for comparable small areas across the UK in maps

and analysis that we would present on our evidence hub website. The small area maps would allow us to discuss in more detail patterns in data related to the wider determinants of health and health inequalities for the whole of the UK.

Would you make use of Census 2021 data for England, Wales and Northern Ireland published alongside each other when they are available, rather than waiting for full UK coverage?

Yes, we will make use of the England, Wales and Northern Ireland data until Scotland data is published.

Section 3: Main changes to variables compared to the 2011 Census

3a: Sexual orientation and gender identity

- New questions on what is your sexual orientation, and is the gender you identity with the same as your sex registered at birth.

For sexual orientation the proposed classification is:

Straight or heterosexual
Gay or lesbian
Bisexual
Other

For gender identity, the proposed classification is:

Yes, gender identity is the same as sex registered at birth
No, gender identity is different to sex registered at birth

In both cases, there would be a further 'not stated' category, which would include any responses that could not be coded and those respondents who did not answer.

Please rank the following factors in order of importance in relation to your needs for sexual orientation and gender identity information.

Smaller geographic detail – 2

Cross tabulated data ie multivariate – 1

Detailed breakdown– 4

Comparable with Scotland and Northern Ireland – 3

We believe we will be able to gain the most insight into social and economic policy issues for this topic through cross tabulation with other variables, such as health, employment, housing. This is the same for all of the topics.

Do they meet our needs?

Yes.

3c: Ethnic group

- In 2011, there were 18 response options in the ethnic group questions. These were divided into five sections. For 2021, they are still developing the most detailed classification.

Which level of ethnic group detail will you use the most?

- **5 high level groups**
- 19 response options
- Full detail of write-in responses

We tend to use higher level ethnic group responses as they are comparable with other survey data sources we use in our analysis. That said, we think it is necessary to see more detail within ethnicity data and would use the 19 response options for relevant analysis.

What geographies will you use the most when analysing ethnic group information?

- National
- Regional
- Local Authority
- **Middle Layer Super Output Area (MSOA)**
- **Lower Layer Super Output Area (LSOA)**
- **Output Area (OA)**

We use ethnicity data at a range of geographies as an organisation. We find that small geographies such as MSOA, LSOA and OA are particularly relevant to analyse ethnicity data. Especially as there can be a lot of variation between small areas, some ethnicities are concentrated in a few MSOAs.

What's the most important priority for us with this data? (user to rank 1-5)

Smaller geographic detail – 1

Cross tabulated data ie multivariate – 2

Detailed breakdown of ethnic group – 3

Comparable with previous censuses – 4

Comparable with Scotland and Northern Ireland – 5

3d: Age, country of birth and other classifications

Age

- With the new capability for users to build their own tables, they've reviewed the classifications needed for each topic. Creating simple groupings of classifications that get progressively more detailed for each topic.

What's the most important priority for us with this data? (user to rank 1-5)

Smaller geographic detail - 1

Cross tabulated data i.e. multivariate - 2

Detailed breakdown of age - 3

Comparable with previous censuses -4

Comparable with Scotland and Northern Ireland – 5

Do our proposals for a reduced set of age classifications meet your needs?

We think that your proposals will generally meet our needs. We think it will be particularly helpful if the flexible data tool gives the opportunity to build bespoke age estimates, as this can often be time consuming.

Disability and unpaid Care

- Disability – more closely aligned with Equality Act 2010 and GSS harmonised questions. The first part of the question asked if the respondent had a condition or illness, and the second part asked if it limited their day-to-day activities.

- Unpaid care – increased the detail for number of hours provided in the range of response options

What would be the impact of including additional detail from any of the variables listed above?

Any additional detail around the number of unpaid carers will allow us to better understand how it impacts people's wellbeing and health. More detailed on the number of hours provided, cross-tabulated with employment and labour market variables, would give us greater insight into how much unpaid care impacts an individual's ability to work.

Splitting the question around disability into two parts, allows us to now identify people that have a long-term condition or illness that does not impact their daily life. This could provide us with much richer insight into the effect of disability on people's wellbeing, labour market engagement, housing situation and more.

Section 4: Proposals for feasibility work to derive new variables

- There is no guarantee that it will be possible to produce these variables, but they are keen to understand user needs better before we investigate them more fully.

4a: Education and employment

8 new proposed variables:

- Resident's route to highest level of education – better understand the qualifications route taken. Eg, are the qualifications gained academic, vocational, or a mixture of both.
- Adult students – by defining them as students in full-time education aged 18 and over and not living with parents.
- Not in employment, education or training (NEETs) – population aged 16–24 years, who are not in employment, education or training.
- Temporarily away from work – because of 'illness, on holiday or temporarily laid off'. This would also include those away from work because of furlough, self-isolation or quarantine.
- Maternity or paternity leave indicator.
- Key or critical worker – people whose occupation was critical in the response to the coronavirus pandemic, as defined by the UK government and the ONS publication on key workers.
- Skills mismatch – where a person's educational level is significantly different to the average level of qualification within their occupation.
- Economic risk created by the coronavirus pandemic – apply the findings of the Business Impacts of COVID Survey (BICS) to census data, to identify populations at financial risk because of the pandemic.

Would you use this variable if it were produced? What sort of analysis might you undertake with it?

We would be interested in the route to highest level of education variable as this data is very difficult to find within other data sources. It would be interesting from an inequalities perspective to understand whether deprivation influences the route to the same qualification.

NEETs data would be used as an outcome variable in our statistical modelling work, primarily to understand whether places with poorer health are more likely to have people categorised as NEET. They would also be interested in looking at the temporarily away from work indicator, to understand its links to health within an area.

We would also use the NEET variable for national benchmarking and for international comparisons. This data could be used alongside the route to education variable to gain more insight around the topic of education.

The key or critical worker variable could be used to understand how this group affected the spread of the pandemic, although a more detailed definition and methodology would be needed to interpret this fully. In other work, we have used survey data to understand the health of key workers, and we will continue to primarily use these sources for analysis even after the Census data is published. That is not to say that the Census variable will not be useful; we would use a key worker variable for insight into the situation in March 2021 and to reweight survey data.

The economic risk variable seems interesting, but it is potentially too backward looking for analysis purposes in our work.

How would you want it to be defined – does the ONS definition work?

The definition for NEETs is suitable. We would like to understand how many people from each age within the defined 16 –24 group are NEET.

We would want a very clear definition of how the key worker group was identified. If possible, we would appreciate a definition that was flexible, and could be adapted or built upon as a user.

What would be the order of importance for this data from the following: (user to rank 1-5)

- Smaller geographic detail – 2
- Cross tabulated data ie multivariate – 1
- Detailed breakdown of variable – 3
- Comparable with Scotland and Northern Ireland – 4

4b: Health and living arrangements

Four new proposed variables:

- COVID-19 health risk – define an output based on the vaccination priority groups by using the self-reported general health status to identify at-risk groups
- Houses in multiple occupation (HMO) – if at least three (unrelated) tenants live there, forming more than one household, and you share toilet, bathroom or kitchen facilities with other tenants
- Multigenerational households – any household with more than two generations resident
- Living apart together – variable indicating the population who live separately to their partners

Would you use this variable if it were produced? What sort of analysis might you undertake with it?

We would be very interested in the households in multiple occupation variable, especially if it could be cross-tabulated with age and care homes.

We believe a COVID-19 health risk would be interesting, but we also ask for a breakdown of long-term conditions and multimorbidity.

We would also use HMO and multigenerational household data as cofounding variables within statistical modelling work. These variables could provide an understanding of to what extent the expected relationship between health and socioeconomic outcomes varies with the introduction of these variables. This would also be possible with the COVID-19 health risk variable.

We would use the multigenerational households to understand the correlation between multigenerational households in areas of low vaccination rates, and impending risks that could arise as a result.

Data on households of multiple occupation would help to build our understanding of housing quality at small geographical detail. Housing quality is a wider determinant of health and therefore is fundamental to our work.

All the variables in this section would help to feed into ongoing work that the foundation is involved in to build an index of social fragmentation within England and Wales.

How would you want it to be defined – does the ONS definition work?

Yes.

What would be the order of importance for this data from the following: (user to rank 1-5)

- Smaller geographic detail - 2
- Cross tabulated data ie multivariate - 1
- Detailed breakdown of variable - 3
- Comparable with Scotland and Northern Ireland - 4

4c: Accommodation type and vacant addresses

Four new proposed variables:

- Care home resident – This separate indicator from communal establishment would denote if a person was resident in a care home.
- Type of vacant address – data from the census collection process which indicates whether a holiday let or second home.
- Transient population – the population who live in a mobile or temporary structure eg a boat or caravan.
- Homeless – including people sleeping rough and sofa surfers.

Would you use this variable if it were produced? What sort of analysis might you undertake with it?

We would use the homelessness and care home resident data as cofounding variables within statistical modelling work.

Teams across the organisation produce analysis on care homes so the care homes variable would be particularly useful to us. This is particularly true as data on social care is very difficult to source. We believe that a care home variable that could be cross-tabulated with other variables would be particularly useful to us.

Homelessness is another variable that we would use if produced, as it is a population we do not have a large amount of information for. It would be useful to compare this variable with statutory data, to understand if there is 'hidden homelessness' which is not shown in official sources. We would be particularly interested in understanding more about 'sofa surfers'. This variable would also inform our analytical work across the foundation, including analysis of the interaction of homelessness with health outcomes.

How would you want it to be defined – does the ONS definition work?

The transient population does not match what one might expect it to mean: someone that moves around frequently with no permanent abode. Someone living in a boat or caravan is likely to be quite fixed.

What would be the order of importance for this data from the following: (user to rank 1-5)

- Smaller geographic detail - 2
- Cross tabulated data ie multivariate - 1
- Detailed breakdown of variable - 4
- Comparable with Scotland and Northern Ireland - 5
- Comparable with previous censuses (if possible) - 3

4d: Additional data needs

Are there any new derived variables you require that have not been discussed in this consultation?

We would be interested in understanding if there is any possibility to derive variables around air quality or environmental exposures based on output areas from the Census.

A broader point around all of these proposed variables is that we would like to understand how we will be able to access them eg as part of the flexible table builder, fixed tables or through prepared analysis. Many of these variables would become significantly more useful to us if they could be cross-tabulated with other data.

Also, we would require clear guidance on the methodology for these variables. This would allow us to evaluate the variables thoroughly and potentially recreate the measures within 2011 data (if ONS is not already planning to do this).

Section 6: Taking a census during a period of change

6a: Place of residence

- In most cases, the coronavirus pandemic will not have impacted the place of residence. However, there will have been some situations where it has done so eg those living at different addresses such as second home or with family/friends, students living at home.

Do you anticipate needing any additional data on place of residence to supplement Census 2021 data?

Some of our work at the Foundation involves triangulation of data to understand how our measures of the population in data like CPRD compare to the Census, so we will continue to do this.

Are there any existing data sources that may meet your needs for place of residence data, either separately or in combination?

Some users within our organisation suggested they might use the population within the NHS digital prevalence estimates admin data to supplement the Census information.

What would be the order of importance for this data from the following:

- Smaller geographic detail - 2
- Cross tabulated data ie multivariate - 1
- Comparable with Scotland and Northern Ireland - 3
- Comparable with previous censuses (if possible) - 4

6b: Economic activity and employment

- The coronavirus pandemic will have impacted many aspects of employment eg furlough, self-isolating, change in characteristics of those looking for work, alternative hours, switching of industries.

Do you anticipate needing any additional data on economic activity and employment to supplement Census 2021 data?

As explained in more detail below we will also use the Labour force survey, Annual population survey and Understanding Society.

Are there any existing data sources that may meet your needs for economic activity and employment data, either separately or in combination?

We would look at the Census information in conjunction with the Labour force survey, Annual population survey and Understanding Society. Census data would help with verification of survey weights, and this alongside the more up to date survey data would provide insight into the labour market.

What would be the order of importance for this data from the following:

- Smaller geographic detail - 2
- Cross tabulated data ie multivariate - 1
- Comparable with Scotland and Northern Ireland - 4
- Comparable with previous censuses (if possible) - 3

6c: Travel to work and workplace zones

- ONS do not know how travel-to-work and workplace-address patterns will evolve following the pandemic. They expect that the transition to greater levels of home working will remain to some extent. However, other aspects such as reduced use of public transport and car sharing may reverse over time.

Do you anticipate needing any additional data on travel to work to supplement Census 2021 data?

It will be difficult to understand the difference between where people work and live past 2021. We appreciate this will be an evolving situation. It would be interesting to see if Census data could be viewed alongside data for travel providers (such as Transport for London) to understand changing commuter patterns.

What would be the order of importance for this data from the following:

- Smaller geographic detail – 2
- Cross tabulated data ie multivariate – 1
- Comparable with Scotland and Northern Ireland – 4
- Comparable with previous censuses (if possible) – 3

Do you use workplace zones?

6d: Impacts on other outputs

- Other potential areas of change include migration, health, disability, and unpaid care.

Do you anticipate needing any additional data to supplement these areas from Census 2021 data?

It would be useful to derive whether someone was shielding from COVID-19 during the time of the Census, as this may give us insight into their behaviour in other areas of life.

Are there any existing data sources that may meet your data needs, either separately or in combination?

We would use the Census data on health alongside Understanding Society, for verification and analysis.

What would be the order of importance for this data from the following:

- Smaller geographic detail - 2
- Cross tabulated data ie multivariate - 1
- Comparable with Scotland and Northern Ireland – 4
- Comparable with previous censuses (if possible) - 3

6e: Emerging data needs

Is there any new analysis of Census 2021 data that you are planning due to the current period of economic and societal change?

We would like to understand from Census data the number of people who were living with their parents during March 2021. This could help us understand its contributions to the spread of the virus. It will also give us an indication of how many people lived in homes that were an unsuitable quality for living in during a pandemic eg those without gardens.

We also believe Census data may help to clarify some of the ongoing issues around understanding vaccine efficacy rates within different age groups.