**Hard to Count index for the 2021 Census**

**Update for the External Assurance Panel**

­­­­­­

Ercilia Dini, MDR/Methodology, Office for National Statistics

May 2019

1. **Background**

The methodology used to develop a Hard to Count (HtC) index for the 2021 Census was presented to members of the External Assurance Panel in October 2018 (Dini, 2018). The members made the following suggestions for further research work to be conducted:

1. MDR/Methodology to consider integrating other data sources into the methodology used to develop the Digital and Willingness domains of the HtC index.
2. ONS to consider social research work to explore reasons for households low/non-response to surveys/census and how this could be improved.
3. MDR/Methodology to investigate international practice and patterns of response to online censuses and consider if these would be applicable to the HtC index for the 2021 Census in England and Wales.
4. ONS to continue working to obtain access to data from other suitable sources (administrative and commercial data).

This paper presents an update on further research work conducted to address the EAP members’ suggestions and our recommendations made based on the results obtained.

Section 2 of this paper presents further research work conducted by MDR/Methodology on both (Willingness and Digital) domains of the HtC index. Section 3 presents ONS work to improve response to surveys and to the 2021 Census. It also presents an update on how information from other countries censuses was considered in the development of the HtC index for the 2021 Census. Section 4 presents an updated on ONS work to obtain access to other data sources. Section 5 presents recommendations made. Section 6 presents the references and Annex A more information on methodology of further research on the domains of the HtC index.

CRAG members are invited to comment on research work conducted and recommendations made.

1. **Further research work conducted on the domains of the HtC index**

The main point highlighted by the members of the EAP was that neither the Digital or the Willingness domain of the HtC index captured willingness to complete a census digitally. In other words, the digital willingness ability was unmeasured. It did not necessarily correlate with either the digital infrastructure of the indicator used to measure households access to broadband internet or to ‘response willingness’ which was modelled on a primarily paper-based census response.

Section 2.1 presents the methodology developed to improve the Digital domain of the HtC index. Section 2.2 presents an update on research work that had been conducted and additional research conducted considering the integration of a digital variable into the Willingness domain of the HtC index.

* 1. **Methodology developed to improve the Digital domain of the HtC index**

A ‘new’ method was developed to improve the digital domain of the HtC index. The method consists of an area level model (LSOA) and uses the Driver and Vehicle Licensing Agency (DVLA) data on online transactions to apply for or renew a driving licence. The methodology used, results obtained by the new method and a comparison of them with the results obtained using the ‘previous’ method is presented in a EAP124.

The model predicts areas according to their digital ‘ability’ to interact with a government website. The model explained 33 per cent of the variance (deviance) suggesting a moderate fit to predict areas digital ability. The predicted values given by the model are used to assign 5 categories of the HtC Digital domain.

At present it is the best proxy to measure people’s interaction with a government website to respond to an online census.

* 1. **Previous and further research work conducted to integrate a ‘digital variable’ into the Willingness domain of the HtC index**

The Willingness domain of the HtC index is constructed using an area level (Lower Super Output) model to build the ***parameters*** to predict response/ non-response to the 2021 Census. The dependent variable is the 2011 Census day 10 non-return rates. The rationale for this is that field follow up started at day 10 after the census day. Returns before that they were therefore self-motivated. The covariates used to build the model parameters are from 2011 Census and updateable administrative data sources. The parameters of the model are applied to up-to-date administrative data. The key assumption in the model is that the structural relationship between the 2011 Census day 10 non-return rate propensity and the key factors remain constant between the 2011 and 2021 censuses despite the change in the mode of collection in the 2021 Census.

We are aware that the ‘Willingness’ domain of the HtC index does not fully capture ‘digital willingness’ but willingness to respond to a primarily paper-based census.

We have considered using the 2011 Census online returns (16% of the households) as the dependent variable to model ‘Willingness’ to respond to a census. In the 2011 Census the online mode of response was optional. Therefore, online returns from the past census were not representative of the whole population in England and Wales and the data is likely to be biased towards the households most willing to do so.

The decision was then to use all up to day 10 returns (paper and online) as the dependent variable to model ‘Willingness’ to respond a census. Despite the ‘Willingness’ domain of the HtC index being built using a conservative approach, the model is robust and based on strong evidence from characteristics of household/people likely respond to past censuses and surveys. Our main assumption is that people who are willing to respond to a census will respond to it whatever the mode of response. And the planning for the census is being made to incorporate several options for those households/people who are not able to respond to an online census. Section 3.2 of this paper provides more information on how this issue has been addressed by ONS.

Before starting the development of the methodology for the Willingness domain of the HtC, we undertook extensive literature review of surveys that collect information on peoples’ use of the internet as an attempt to build a ‘combined’ willingness-digital index. We reviewed the information collected by ONS-Labour Force Survey Quarterly updates (ONS, 2014), ONS-Opinions Life Survey (ONS, 2014-2018), Cabinet Office-Digital Landscape survey (CO, 2012), Oxford Internet Surveys (Oxford Internet Institute, 2013), Ofcom’s Media literacy research (Ofcom, 2015) and CDRC geodemographic of internet use and engagement (CDRC, 2014). Where information was available we conducted analysis by broad age groups and sex. In line with the results from the 2017 Census test (Corps et al, 2017), younger age groups are more likely to use internet and to use the internet to interact with government websites. Despite the results being good indicators of internet use by younger age groups we have to be cautious about making inferences to the whole population (of both young and other age groups). The surveys are voluntary and likely to have response bias (including the 2017 Census test). The pattern of response may be different for a census which is mandatory. Also, the results from surveys provided information at national level but sample sizes were not large enough to give accurate estimates at LSOA level of geography. We would need a complex method to combine the 2011 Census and up-to-date survey’s information to the level of geography needed. And this method would still have the risk of inaccuracy of results.

We tested using the results of the 2017 Census test to build an ‘online adjustment’ to the Willingness domain of the HtC index. Considering the small sample size, the low response rates and the risk of response bias in the 2017 Census test, we decided not to pursue this method. First because the 2017 Census test did not collect information designed to be used as an adjustment variable to the willingness to respond to a census model. Second, even using other information that had been collected there would be a high risk of incurring in an error that could result in spoiling a robust model which, despite being conservative, is based on results from a census and strong evidence of characteristics of people who are willing to respond to it.

We also tested the use of Principal Component Analysis (PCA) (Congdon, 2003) to combine the variables, 2011 Census day 10 returns and access to the internet (Ofcom data), to build a ‘Willingness-digital’ index. It did not prove to be an appropriate method. PCA is a variables’ reduction technique**[[1]](#footnote-1)** and used when primarily all variables are measuring the same concepts and there is evidence that there is ‘redundancy’ in some of the variables. In our case the variables (census return and digital access) measure different concepts. The results obtained by combining these variables using the PCA method did not give us a clear outcome of what was being measured. The same principle applies if we include the DVLA data as another variable and conduct a PCA.

Further research work was conducted to test the incorporation of DVLA data online transactions to apply for or renew a driving licence as a covariate in the Willingness model (to predict response/non-response by day 10 after census day). The results did not show an improvement to the previous model. Annex A presents the parameters of the model and a comparison with the parameters given by the previous model.

1. **ONS work to increase responses to surveys and to the 2021 Census**

Section 3.1 presents an update on ONS recent work to increase response rates to surveys and Section 3.2 presents an update on ONS strategy to increase responses to the 2021 Census. It also provides an update on how international practice has been taken into consideration in the development of the HtC index.

* 1. **ONS work to increase response rates to surveys**

ONS has been investing in transforming its surveys and is not alone in its pursuit of adding online data collection to its portfolio. This means that our surveys will be online first, with traditional use for follow-up. The surveys are being transformed to ensure that data collection is more efficient, and the effort required by the person responding is reduced.

A more ‘User- Centred Design’ is being developed. Introducing self-completion means that we can no longer rely on highly trained interviewers to achieve response and to provide a good experience. To address this, ONS is fundamentally changing the design of its surveys. Below is a summary of some of the changes being made to ONS surveys in an attempt to address the shift to the online collection and increase response rates.

* Designing mobile first, ensuring surveys work on smartphones in the first instance. Constraining content to less screen space to make questions clear to respondents;
* Optimising questionnaire content;
* Looking at how a respondent thinks, and how they conceptualise a topic. This has resulted in reordering the questionnaire content to help respondents;
* Learning from interviewers and seeking their assurance on our prototypes; our transformation starts with them. We speak with them first to learn about the current problems with the wording and flow and look to address those in our redesigns;
* Using a conversational tone and using words that respondents would understand. We are also writing for the average reading age of the UK public;
* Using behavioural insights to design respondent materials. We only include content that is needed by a respondent. Our testing has revealed that less is more (e.g. shorter materials are better received by users).

ONS is also heavily investing in its corporate services and infrastructure to deliver this transformation. This includes integrating our systems and sharing them across business areas – employing the principle of ‘build the thing once’. More information on how ONS is transforming its surveys is available in an ONS recent publication (Wilson, 2018).

* 1. **ONS strategy to increase responses to the 2021 Census**

ONS has investigated responses to the latest online censuses conducted in Australia (Australian Bureau of Statistics, 2016) and Canada (Statistics Canada, 2016). It also investigated the United States annual planning database of indicators used to estimate areas/groups with low response scores (LRS) which measures the perceived difficulty of counting groups and tract levels based on mailed-in self-response. Of the 25 LRS predictor variables, percentage of renters in certain groups, percentage of people aged 18-24, and percentage of households headed by unmarried females are the highest predictors of low response rates (US Bureau of Statistics, 2018).

The results from the 2016 Australian census showed that 59 per cent of the responses from occupied private dwellings were received online in 2016, up from 34 per cent in 2011. There were small differences in online response rates between ages 0 – 49 years. The online response rate reduced in older age groups. However, 52 per cent of people aged 60-69 years participated online and more than one in four of those aged 80 years and over (28 per cent) were counted in an online form. People born outside Australia were more likely to participate online (69 per cent for those in private dwellings) than people born in Australia (61 per cent). People who spoke a language other than English and could not speak English at all were more likely to complete the online form (83 per cent) than people who only spoke English at home (60 per cent).

We considered if their experience could be applicable to the development of the HtC index for the 2021 Census in England and Wales. Information from the Australian census and the planning database of indicators from the US Bureau of Statistics were useful to confirm that low responses usually occur in population groups with particular characteristics. Despite historical and cultural differences between countries, we share most of the census non-respondent characteristics. And most of these characteristics were included in the model to build the Willingness domain of the HtC index for the 2021 Census in England and Wales.

ONS has planned a strategy to maximise response in the 2021 Census by using a combination of the two domains of the HtC index. These will be used to plan where paper questionnaires will be sent as first contact to households and where assisted digital centres should be situated.

Questionnaires will be available via a number of routes to ensure that, although we will encourage anyone who is able to respond online to do so, there are no barriers to completion for those who cannot, or choose not to, complete their forms online.

There will be four different options for receiving paper questionnaires:

Paper as first contact

In some areas of England and Wales, where the digital ability is likely to be low, the initial contact will be a paper questionnaire, delivered by post. These paper questionnaires will also include a unique access code (UAC) to enable those who prefer to complete the census online to do so.

Paper questionnaire as a reminder

In some other areas, where the initial contact letter includes only an UAC for online completion, a paper questionnaire will be posted out at a later date to non-responding households as a reminder to respond.

Via contact centre or online

Anyone who wants to complete a paper questionnaire will also be able to request one at any time either via the website, or by calling our helpline, and information about how to do this will be included in the initial contact letter inviting people to respond.

Handed out by field officers

Census field officers will visit all addresses where no response has been received to remind people to respond. At this point, field officers will also be able to supply paper questionnaires on the doorstep if requested.

Other forms of assistance

For those who do wish to complete the census questionnaire online, but may require help to do so, face-to-face assistance will be offered through an assisted digital service at a variety of locations. ONS will also be providing help with completion at organised completion events and help will also be available via our contact centre.

1. **ONS work to obtain access to other data sources**

A major milestone to ONS for obtaining access to data sources has been the ‘Digital Economy Act’ in 2017. The Act clears the numerous legal barriers that have evolved over time preventing us getting the data we need, when we need it. The Act gives us legal rights to access data held by Government Departments, other public bodies, charities and large/medium-sized businesses where we need that data to support our work.

Negotiations between ONS and other Government Departments to agree terms of data sharing have been held for some time and ONS already has access to a number of administrative and some commercial data sources.

ONS is in continuous negotiations to ensure we have access to data needed to produce better statistics. More specifically, for the development of the HtC index for the 2021 Census we would ideally have access to Electoral Register data at LSOA level of geography. At present the agreement between ONS and the Cabinet Office is for them to provide us with data at Local Authority level. ONS is considering to re-open the negotiation to feed the requirements of access to the data at the level of geography needed to further improve the Digital domain of the HtC index.

ONS is not only gaining access to administrative data sources. It is also conducting research to investigate the statistical quality of the data and the scope of their use for either producing Research Outputs or to enhance 2021 Census outputs. Information on new data sources investigated for the Administrative Data Census Research Outputs, as well as data sources being assessed for use in the 2021 Census have been published and are available at ONS website (ONS, 2019).

**Recommendations**

The recommendations presented in this section refer to research work conducted and presented in sections 2.2, 3.2 and 4 of this paper. The recommendations made referring to the methodology developed to improve the Digital domain of the HtC index (Section 2.1 of this paper) are included in that paper.

* 1. Recommendations referring to Section 2.2
* We do not recommend the inclusion of DVLA data on online transactions as a covariate into the model used to build the Willingness domain of the HtC index for the following reasons:
1. we do not have DVLA data for 2011 to build the parameters for the model and apply them to more up-to-date DVLA data to get the predicted values. DVLA data provided to ONS goes back to 2016. Using 2016 data to build parameters for a 2011 outcome variable would not be best methodological approach.
2. the results given by the model including DVLA as a covariate (despite the constraints explained in item a) did not improve the fit of the previous model to a significant degree to justify its inclusion.

The predicted values given by the model to build the Willingness domain of the HtC index should not be taken as the actual responses to the 2021 Census. The predicted values are to be used only as the likelihood of LSOAs (a rank of areas) according to their willingness to respond to a census.

The HtC categories of the Willingness domain should be used in conjunction with the HtC categories of the Digital domain for planning areas where paper questionnaires are to be sent as first approach to households.

* 1. Recommendations referring to Section 3.2

We highly recommend the proposed plan for a strategy to maximise response in the 2021 Census using a combination of the two domains of the HtC index. The HtC domains will be used to plan where paper questionnaires will be sent as first contact to households and where assisted digital centres should be situated.

The four different options for receiving paper questionnaires will be tested in the 2019 Census rehearsal. We recommend the results to be analysed to evaluate if both domains of the HtC index are good discriminators of online and paper responses.

* 1. Recommendations referring to Section 4

ONS is in continuous negotiations to ensure access to data needed to produce better statistics.

We recommend ONS to continue the negotiations to gain access to the Electoral Register data at LSOA level of geography. This should be considered according to ONS priorities for access to other data sources relevant to the 2021 Census.

**References**

Australian Bureau of Statistics (2016). Note on information provided by the Australian Bureau on results of the 2016 Census. ‘There was little difference in online response rates between ages 0 – 49 years. The online response rate reduced in older age groups. However, the majority of people aged 60-69 years (52.2%) participated online and more than one in four of those aged 80 years and over (28.3%) were counted in an online form. This age group’s online participation increased by the greatest proportion in 2016 with the online response rate increasing 225%. People born outside Australia were more likely to participate online (68.6% for those in private dwellings) than people born in Australia (61.0%). People who only spoke English at home were the least likely to participate online (59.9%), whereas people who spoke a language other than English and could not speak English at all were the most likely to complete the online form (82.9%)’. More information is available at:

[https://www.abs.gov.au/websitedbs/d3310114.nsf/home/Independent+Assurance+Panel/$File/CIAP+Report+on+the+quality+of+2016+Census+data.pdf](https://www.abs.gov.au/websitedbs/d3310114.nsf/home/Independent%2BAssurance%2BPanel/%24File/CIAP%2BReport%2Bon%2Bthe%2Bquality%2Bof%2B2016%2BCensus%2Bdata.pdf)

<http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20Census%20online%20responses%20infographic>

Cabinet Office (2012). Digital Landscape Research. Available at: <https://www.gov.uk/government/publications/digital-landscape-research/digital-landscape-research>

Congdon P (2003). Bayesian Statistical modelling. John Wiley & Sons, Ltd, p. 394-410.

Consumer Data Research Centre (CDRC, 2014). Internet user classification 2014. Available at:

<https://maps.cdrc.ac.uk/#/geodemographics/iuc14/default/BTTTFFT/10/-0.1500/51.5200/>

Corps D, Fraser O and Meirinhos V (ONS, 2017). Test Analysis – Analysis of online and paper respondents. ONS Internal report. Available at: <https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD_Stats_Design/Statistical_Design/2017%20Test/Reports/CRAG_paper_Nov17.docx>

Digital Economy Act (2017). Available at:

<https://www.gov.uk/government/collections/digital-economy-bill-2016>

Dini E (2018). Hard to Count index for the 2021 Census. ONS Internal report. Available at:

[https://share.sp.ons.statistics.gov.uk/sites/MTH/Cen/2021/Sample\_Design\_and\_Estimation/Hard\_to\_Count/HtC\_index\_report\_v5\_for\_External\_Assurance\_Panel.docx](https://share.sp.ons.statistics.gov.uk/sites/MTH/Cen/2021/Sample_Design_and_Estimation/Hard_to_Count/Census_2021_HtC_index_report_v5_for_EAP.docx)

Dini E (2019). Update on methodology for the Digital domain of the Hard to Count index for the 2021 Census. ONS Internal report. Available at:

<https://share.sp.ons.statistics.gov.uk/sites/MTH/Cen/2021/Sample_Design_and_Estimation/Hard_to_Count/Update_on_methodology_for_the_Digital_domain_of_the_Hard_to_Count_index_for_the_2021_Census_May_2019.pdf>

Ofcom (2015). Ofcom’s Media literacy research. Available at <https://www.ofcom.org.uk/research-and-data/media-literacy-research/adults/adults-media-use-and-attitudes>

Office for National Statistics (2014). Statistical bulletin: Internet Access Quarterly Update, Q1 2015. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccessquarterlyupdate/2014-08-20>

Office for National Statistics (2018). Statistical Bulletin: Internet Access – Households and Individuals 2018. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018>

Office for National Statistics (2019). Administrative data census project. Available at:

<https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject>

Wilson L (2018). Survey Methodology Bulletin. v 78, p. 45-52. <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/surveymethodologybulletin>

A summary is available at: <https://blog.ons.gov.uk/2019/01/10/designing-future-surveys/>

Oxford Internet Institute. University of Oxford (2013). The Oxford Internet Surveys. Available at: <http://oxis.oii.ox.ac.uk/>

Statistics Canada (2016). 2016 Census of Population collection response rates. Results by region. Information available at:

<https://www12.statcan.gc.ca/census-recensement/2016/ref/response-rates-eng.cfm>

U S Bureau of Statistics (2018). Mapping Hard to Count (HTC) Communities for a Fair and Accurate 2020 Census. Information available at: <https://www.censushardtocountmaps2020.us/>

<https://datadrivendetroit.org/blog/2018/02/13/measuring-hard-to-count-populations/>

**Annex A**

Table A1 presents the parameters given by the Willingness model as it was presented initially to the members of the EAP (i.e. the model does not include DVLA data on online transactions as a covariate). Table A2 presents the parameters given by the Willingness model including the DVLA data as a covariate to the model.



The parameters of the model that includes DVLA data on online transactions as a covariate (Table A2) do not show large differences when compared to the parameters of the previous model (Table A1). Both models explained 70 per cent of the variance (deviance). In other words, both models ‘explained’ 70 per cent of the variation in the total LSOAs’ household response/non-response observed in 2011.

Table A3: Number of LSOAs in HtC categories 1 to 5 given by the Willingness model with or without the use of DVLA data as a covariate in the model. England and Wales.



Table A4: Percentage of LSOAs’ agreement in HtC categories 1 to 5 given by the Willingness model with or without the use of DVLA data as a covariate in the model. England and Wales.



Table A5: Kappa statistical test of agreement in HtC categories attributed to LSOAs by each model. LSOAs in England and Wales.



 \* Kappa value = 0.82 = Almost perfect agreement (Landis and Koch, 1977)**[[2]](#footnote-2)**

Table A6: Number of LSOAs in HtC categories 1 to 5 given by the Willingness model with and without the use of DVLA data as a covariate in the model. LSOAs in Inner and Outer London.



Table A7: Percentage of LSOAs’ agreement in HtC categories 1 to 5 given by the Willingness model with or without the use of DVLA data as a covariate in the model. LSOAs in Inner and Outer London.



Table A8: Kappa statistical test of agreement in HtC categories attributed to LSOAs by each model. LSOAs in Inner and Outer London.



 \* Kappa value = 0.81 = Almost perfect agreement (Landis and Koch, 1977)

The lowest agreement was seen for category 1 of the HtC. This can be partially explained because very few LSOAs fall in that category. This is the case independently of the model used. As the Kappa statistic test is a weighted test and category 1 of the HtC had the lowest number of observations in London, the overall agreement is still substantial (0.81) (Table A8).

1. ‘Principal component analysis is a variable reduction procedure. It is useful when you have obtained data on a number of variables (possibly a large number of variables) and believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, you believe that it should be possible to reduce the observed variables into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables.’ [↑](#footnote-ref-1)
2. Landis, R and Kock, GG (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, Vol. 33, No. 1 (Mar 1977), pp. 159 [↑](#footnote-ref-2)