**Independent Methodological Review: Response Chasing Algorithm**

# Summary

The 2021 Census will be predominantly online for the first time. As a result, we will have a lot more information much more quickly than ever before on who has submitted a Census questionnaire. Being able to use this information intelligently during live operations enables us to maximise the effectiveness and efficiency of our operation by tailoring interventions to the areas and populations where they are most needed.

The Response Chasing Algorithm (RCA) is an innovative tool developed by the Census Statistical Design (Collection) team that enables live return rates to be compared with expected patterns, identifies and prioritises any shortfalls in return rates and recommends appropriate interventions to address the issues. In a world where it is increasingly hard to get people to respond to any survey or census, having this sort of tool to aid decision making and ensure we make the best use of resources to target problem areas is essential to ensure that we meet our success criteria and deliver high quality census estimates.

A combination between business and statistically designed indicators is expected to provide constantly updated information required for a timely and well-informed decision-making process. It is anticipated that this tool that will be able to facilitate timely and well-informed decisions during the decisive live collection period. The Response Chasing Algorithm has been reviewed by the Census Transformation Programme and international experts from other Censuses (e.g. Canada, Australia and New Zealand).

**Table of contents**

Table of Contents

[Summary 1](#_Toc10633441)

[1. Introduction 3](#_Toc10633442)

[2. Method 3](#_Toc10633443)

[2.1. Hard to Count index 3](#_Toc10633444)

[2.2. Self-response Profiles 5](#_Toc10633445)

[2.3. Field Operations Simulation 6](#_Toc10633446)

[2.4. Forecast model 8](#_Toc10633447)

[3. Interventions 10](#_Toc10633448)

[3.1. Field Operations 10](#_Toc10633449)

[3.2. Reminder letters and paper questionnaires 11](#_Toc10633450)

[3.3. Assisted Digital 11](#_Toc10633451)

[3.4. Media campaigns 12](#_Toc10633452)

[4. Interventions table 13](#_Toc10633453)

[5. Visualisation Tool 14](#_Toc10633454)

[6. Discussion 16](#_Toc10633455)

[7. References 18](#_Toc10633456)

[8. Appendix 19](#_Toc10633457)

[8.1. RCA Prioritisation Algorithm - Rules and Processes 19](#_Toc10633458)

# Introduction

The Response Chasing Algorithm (RCA) is a product developed by the Statistical Design (Collection) team which aims to support the Census Operations during the collection period. It has been developed to analyse daily Census returns and responses, in order to provide insight about non-respondents and to propose maximising response recommendations (Ardis, Shipsey, & Meirinhos, 2018). During live operations, it is anticipated that this tool will be used by the Census Operations and different business teams and therefore it is expected that this tool will contribute to achieve the desired Census Targets (i.e. 94% overall Census response, 75% online response, at least 80% response across local authority, and low variability).

As a business decision support tool, the RCA has also been designed to maximise the impact of the analytical and visualisation packages (i.e. algorithms, forecast and dashboard). When all these components are combined and available, users are provided with reliable information about census return and response patterns, and different measures of the statistical quality of the census data.

The RCA statistical package has 4 distinct components:

1. RCA Prioritisation Algorithm;
2. RCA Forecast Model;
3. RCA Dashboard;
4. Field Visits Prioritisation Algorithm.

This review will focus on the first 3 components, with the Field Prioritisation Algorithm being covered in a separate report (Meirinhos, 2019a).

# Method

In order to make recommendations for effective interventions, in addition to analysing responses and returns from the census, there is a need to understand the data better and take into account aspects such as the Hard to Count (HtC) index, forecast models of expected returns and responses, demographic information (such as age and sex). To identify and correct problems, the algorithm examines daily target data (modelled by the Field Operations Simulation - FOS) with live returns from online and paper Census submissions.

##  Hard to Count index

The Hard to Count index has been first developed for the 2001 Census when there was evidence that non-response was correlated with certain household and individual characteristics. Based on these, the lower-layer super output areas (LSOAs) in England and Wales have been split into one of 5 categories, with 1 being the easiest to count, and 5 being the hardest to count areas, illustrating people’s willingness to respond[[1]](#footnote-1).

However, as the primary collection mode for the 2021 Census will be through an online questionnaire, which is different to previously conducted censuses, the non-response patterns observed before might also be different. In addition to the willingness hard to count part of the index, there was a need to develop a digital one, based on people’s access and use of technology. This domain is still in development as new sources of information become available.

Both parts of the index split the LSOAs into 5 categories as follows: 40% in HtC 1, 40% in HtC 2, 10% in HtC 3, 8% in HtC 4 and 2% in HtC 5 (Dini, 2018). The process of combining both indexes has been used to design the collection strategy for the 2021 Census.

|  |  |
| --- | --- |
|  | Willingness |
| 1 | 2 | 3 | 4 | 5 |
| Digital | 1 | 4,853,51118.70% | 3,310,75712.76% | 727,0502.80% | 1,101,6074.25% | 369,2841.42% |
| 2 | 4,206,29016.21% | 4,388,76316.91% | 1,006,6783.88% | 640,0052.47% | 136,0060.52% |
| 3 | 769,3802.97% | 1,317,6145.08% | 353,2371.36% | 156,8330.60% | 23,4460.09% |
| 4 | 396,7321.53% | 1,086,5174.19% | 369,7241.42% | 187,5590.72% | 34,8720.13% |
| 5 | 57,8980.22% | 231,2750.89% | 102,5230.40% | 108,6000.42% | 12,5390.05% |

Paper questionnaire as a reminder

Reminder letter

Paper questionnaires as initial contact

Figure 1. Table showing the paper strategy by HtC willingness and HtC digital with provisional numbers for 2021 Census

As shown in the figure above, the Hard to Count index feeds into the strategy for sending paper questionnaires and reminder letters (further discussed in section 3.2.)

##  Self-response Profiles

Response profiles are a best estimate of the volume of responses we expect to receive by geographical areas during the census period. They model when, where and how people are going to respond to the census, and they will be vital to inform operational decisions during the 2019 Rehearsal and the 2021 Census (Meyer, Mok, & Sullivan, 2015; Salganik, 2017).

In advance to the 2019 Census Rehearsal and the 2021 Census, updated response profiles will feed into the Field Operation Simulation Model (FOS) which will be used to help make decisions about the number of field staff required in each area, where to send reminder letters, and where to send paper questionnaires as the initial contact. During the live collection period, the simulated data derived from the response profiles and including all planned collection strategies will be used by the Response Chasing Algorithm tool (RCA) as a basis for comparison against live return data, so that we can tell where we need to put in extra effort to meet our quality targets of 94% overall response and at least 80% response in every local authority.

The response profiles model provides information and tolerances around the expected number of returns by day, from the start of the census period until the end of the collection period for each mode of response – online or paper. Our objective is that each lower super output area (LSOA) will have a response profile which is associated with its influential demographic variables and Hard-to-Count (HtC) rating. Groups of LSOAs that have similar self-response profile shapes and demographics, will be clustered together.

We anticipate that a response profile might look like the graph in figure 2. The blue line shows the overall expected response. This is broken down to show the daily expected online (grey line) and paper (orange line) responses.

Figure 2. Projected return profile for the 2021 Census showing different response peaks (day 7 = initial contact lands; day 21= Census day; day 28 = reminder letter 1).

## Field Operations Simulation

The 2021 Census collection operation involves many resources and high costs. To allocate these resources efficiently and save money while ensuring we meet our response targets, there is a need to cheaply simulate different scenarios and study the impact of changes in our collection operation strategy.

Discrete Event Simulation (DES) is a computer-based modelling method that can be used to represent complex but well-defined systems by breaking them down into small, manageable parts. The Census collection operation is a good example of such a system, therefore a DES was developed called the Field Operations Simulation (FOS).

The FOS was created using Python to take advantage of its free, open-source nature, as well as a large and active community of Python users both within the analytical function of ONS and more widely. Despite its name, the FOS goes beyond field operations into many other aspects of the census collection operation. Here are some example questions it can answer:

* In order to send a reminder letter to every non-responding household, how many do we need to print, and when?
* To which areas should we send a paper questionnaire as a first point of contact, and what impact will this have?
* How many extra field staff must we employ to counter the effect of a decrease in self-response?

To validate the simulation, staff from across the programme were invited to attend workshops where they could set their own scenarios to see how the simulation behaved. This was subsequently validated using 2011 Census data and data from the 2017 test (Barber, 2018). Further validation and updates will be performed after the 2019 Rehearsal. In general, there is a good agreement between reality and the simulation. Where discrepancies exist, these can be explained by either operational issues during the test or the differences between a test and a census (e.g., one is compulsory, and one is not). However, the simulation outputs are entirely dependent on the choice of input parameters, and there is unavoidable uncertainty when modelling future events for which no closely related real-world scenarios are available[[2]](#footnote-2).

The FOS is highly dependent on other epics such as the Hard to Count index and the self-response profiles (estimated rates of self-response by day by local area generated by time series analysis). Based on these, the simulation creates simulated response profiles, which represent the best estimate of the volume of responses we expect to receive during the census period. The results feed into the RCA to be used during live operations to find out where extra efforts are needed to reach our targets. They are also used more generally across the organisation in determining the numbers to be used for procurement, for example the number of field staff required in each area, how many reminder letters and paper questionnaires are needed and where to send them as an initial contact.

The aim is to build simulated response profiles for returns by day, for the entire collection period and for each mode (online and paper). The objective is to assign a profile to each LSOA, associated with its influential demographic variables and HtC rating. Groups of LSOAs that have similar self-response profile shapes and demographics, will be clustered together. This means that we can deal with the smallest possible number of response profiles, rather than 34,753 (the number of LSOAs).

This approach is then used to build up the return profiles for all desired levels of geography (Local Authority, Regional or National).

Figure 3. Expected cumulative returns for the 2021 Census by Hard-to-Count classification. It is expected that areas within HtC 1 will be more likely to self-respond and HtC 5 will be less likely. Extra effort will be required to maximise response in HtC 4 and 5 areas.

A challenge in building the 2021 Census response profiles is that in 2021 we are moving to an ‘online-first’ strategy. This means that online engagement will be actively sought and prioritised ahead of non-online modes. This objective makes the task of predicting responses for the 2019 Rehearsal and 2021 Census more difficult. Comparisons to the 2011 Census or other paper-based censuses are limited due to this fundamental change in approach. A possible approach to develop the response profiles considers not only the willingness of the public to respond to a census, but also their willingness to respond online.

An additional challenge to our ambitions comes from the fact that we know that sending a unique access code (UAC) and invitation to complete the census online, does have a negative impact on response when compared with sending a paper questionnaire. This was shown in the 2017 Test when the sample who were sent paper questionnaires had a response rate of 39% compared with the 26% response rate achieved by the comparative sample who were sent invitations to complete online.

However, of the households who were sent paper questionnaires first in the test, 30% of the responding households did respond online. Comparing this with the 16% online submission rate for the 2011 Census, we can see that people are becoming more willing to complete the census questionnaire online. Evidence from the 2017 Test also shows that people tend to respond in the mode first offered to them. We therefore do not want to provide too much paper in the first instance but would rather predict where this is the most effective approach so that we can achieve the quality target of 75% online response.

In addition, we must bear in mind the need to control variance in response – keeping variation low whilst simultaneously seeking high response rates. A relatively high response rate is needed to prevent a substantial increase in variance when adjusting for non-response (Racinskij, 2015).

To achieve the 75% online response target, whilst pursuing the highest response rate and low variance, Census Operations might be using the response profiles as a source of consistent information about the underlying characteristics of the population for each geographical area (LSOA or LA) during and before the census collection period. Evidence from the 2016 Test[[3]](#footnote-3) regarding the preferred mode of census completion by age distribution for online respondents suggests that there should be a focus on the age group 75 and over to achieve the desired target for this segment of the population (when asked about the preferred mode of response to census only 48% of the sample in this group mentioned online response as their choice to reply to the census).

## Forecast model

During 2021 Census, the Operations Team will deploy a variety of interventions (e.g. field visits, sending reminder postcards and questionnaires, media campaigns) to make sure we achieve our target return rates. The 2019 Rehearsal aims to test the effectiveness of those interventions and of our operational processes.

The Response Chasing Algorithm (RCA) will also help with achieving the desired response rates. Its main role is to flag shortfalls in response rates in specified geographical areas, and to propose interventions to ensure that we reach our targets. In more detail, the algorithm will receive daily live response rates and will compare them with what we expect to achieve based on previous Censuses and on the results of the Field Operation Simulation model. The basic geography level for which recommendations will be produced is LSOA, however shortfalls will also be shown for all other geography levels.

During the census period, the RCA will not only show current and expected response rates for the given day, but it will also show the forecasted response rates until the end of the census collection period. Benefits of the forecasting method are three-fold:

a) it shows what we can expect to achieve at the end of Census collection if we don’t intervene;

b) it can be used to justify the proposed interventions in a visual and tangible way - people in charge can understand what will (possibly) happen without interventions;

c) they enable us to prioritise areas that need interventions.

To calculate the forecasted response rate values for the given LSOA, we will work with non-response rates. This is to ensure the forecasted values are always going to fall below 100%. If we were working with response, as opposed to non-response rates, final predicted response rate could be above 100%, which should not happen.

The equation used is as follows:

$$\frac{FN}{ON}= \frac{TN}{EN}$$

$$FN=ON × \frac{TN}{EN}$$

ON = observed non-response rate (e.g. in day 20)
EN = expected non-response rate (e.g. in day 20)
TN = target non-response (e.g. after all operations concluded)
FN = forecasted non-response

(By using this equation, we are assuming that FOS is accurately predicting the percentage chance of each non-responding HH to respond in the future.)

The graph below shows which numbers are used to calculate forecasted non-response. In this example, the forecast is calculated for the final response day. The same formula has to be used for each day to produce the green dotted line.



Figure 4. A forecast for an LSOA in Ceredigion during the 2021 Census using simulated “live” data.

# Interventions

The main role of the RCA is to flag shortfalls in response rates in specified geographical areas and to propose interventions to ensure that we are on target to reach the desired response rates. Interventions can be related to field staff, reminders, assisted digital centers or media campaigns.

## Field Operations

Field staff are an integral part of the Census Operations as they help us achieve our targets but are also an expensive component, so we want to make the best possible use of them. Field follow-up begins two days after Census Day, where a percentage of staff focus heavily on HtC 4 and 5 areas. The rest of the field goes out nine days after Census Day, and that is when the RCA can start recommending interventions that relate to field follow-up:

* Increasing field staff hours;
* Moving field staff from one Coordinator Area to a neighbouring one;
* Moving field staff nationally (1% mobile workforce).

With all these interventions, there are increased costs that need to be considered together with the potential effectiveness, to assess which one is most appropriate to recommend. With increasing field staff hours, the costs would be associated with salaries and mileage, but when moving field staff in different areas, there is the potential of having to increase their pay because different Local Authorities (LAs) are in different pay bands.

##  Reminder letters and paper questionnaires

When it comes to print, there are two types of interventions: planned and unplanned. The planned interventions are set up well in advance and independent of the RCA (we know when and where we will send reminder letters and paper questionnaires, as well as how many of each will be sent out). The unplanned interventions, while they have some planned aspects (such as when they will be sent out), how many and where will be dependent on the RCA’s recommendations.

Reminder letters proposed by the RCA (i.e. response driven – RD) will be sent to areas where the return rate is substantially below the expected targets. There are some operational constraints associated with RD Reminders such as the 5-day lag between cut & delivery and the fact that they will be sent in 3 or 4 batches between the planned reminder letter batches. This means that households getting a planned reminder in the next batch would be excluded and that no household would receive more than two (planned or unplanned) reminders. A similar approach will be used for interventions with paper questionnaires.

## Assisted Digital

The main purpose of the Assisted Digital (AD) centres is to offer support to those who have limited digital skills but would still like to complete the census questionnaire online. The experience of the recent New Zealand census demonstrated that face to face help is an essential element of success. If it is not implemented as a dedicated service, this may affect the quality of the data collected. This has emerged as a factor in New Zealand’s unsatisfactory result, reporting that people such as the elderly, those in rural communities and those with disabilities faced greater problems participating in the 2018 Census.

Based on the findings from the Assisted Digital pilot for the 2017 Census Test and research into segments of the population lacking in digital skills, it is reasonable to assume that at least 2 per cent of those completing the census will require face to face help. That equates to 500,000 responses. Because the test was voluntary and census participation is compulsory, we can anticipate a significantly higher figure. If, for example, 5 per cent of respondents want help, the figure would rise to 1.2 million.

Research using the HtC index and Ofcom 2016 regional internet availability figures means we can look at and combine high or low availability with those groups showing the highest degree of willingness to complete. This should give us our most accurate indication of where we need to offer assisted digital services.

Where the centres will be set up is dependent on where the RCA flags the problems during live operations, and also on previously known information such as demographic breakdown of the areas and the digital domain of the HtC index (Bexley, 2018). The scope of the service delivery includes areas with:

* High to moderate digital availability;
* Significantly high degrees of willingness to complete;
* Significant deprivation.

It excludes areas that receive paper first, as research shows that people are more likely to complete the questionnaire on paper in that case.

## Media campaigns

The aim of media campaigns is to reach the varied audience of 26 million households effectively, and in order to achieve that a mixture of direct channels and partnership led activities will be used (ONS, 2019).

The media channels are categorised into three groups:

1. Paid media – used to drive audiences to our “owned” media channels:
	* Broadcast (TV, radio, cinema, outdoor display, print advertising, sponsorship) – the role of broadcast is mass awareness and activating response;
	* Advertising;
	* Paid search;
	* Paid social media (The audience for paid media will be those who are unaware of the census/ONS and therefore we need to reach out to).
2. Owned media, primarily digital, which is hosted on platforms populated by our audiences (such as Facebook, Twitter, YouTube etc), signposted to our channels by paid media. Once here, we need to engage and retain the interest of our audiences with strong and relevant content.
3. Earned media offers free publicity generated by audiences positively engaged by our content and by advocates:
	* Advocates’ social channels – these platforms are crucial in reaching beyond our paid advertising into areas where the census messaging might not ordinarily penetrate;
	* Reposts;
	* Shares;
	* News articles and reports.

Positive earned media has the greatest potential to grow profile, popularity and audience involvement in the census and ONS.



Figure 5. Timeline for the 2019 Rehearsal with the different stages from the ONS media campaign and the messages to be highlighted during each stage (in dark pink)

# Interventions table

The analytical features of the RCA Analytics (prioritisation, intervention and forecast algorithms) have been refined to provide users with up-to-date operational insight for each one of the maximising response strategies (RCA interventions: media campaigns; assisted digital; reminder letters, paper questionnaires and increased field staff hours).

The intervention selection process is designed to provide interventions specifically customised to each LSOA by considering the relative impact from each strategy (modelled from the FOS) as well as anticipated operational constrains (business indicators: availability, delay/lag and cost). To achieve this requirement, statistical and operational information is used by the intervention algorithm to provide one or combined maximising response recommendation(s) to all LSOAs flagged as risk areas by the prioritisation algorithm (low response and/or high variability).

The selection process starts with all interventions being considered as possible solutions for response or variability issues. All interventions (ranked by higher impact in response) move throughout the algorithm pipeline (rule set). Once this starts to happen, interventions are classified as trials. When a given trial stops its journey because it does not fulfil the next rule, it is then qualified as blocked. If one specific trial progresses until the end of the rule set it becomes a recommendation.

The process continues until all interventions are classified as blocked or recommendations. If the combined recommendations’ impact is higher than the gap in response to be resolved, then the top ranked recommendations are used (up to the value to be solved). This system ensures that the RCA prioritisation algorithm is able to consider concurrent recommendations, especially in areas where one alone is not enough to increase the response to the desired levels. But also allows new recommendations to be considered when new operational constrains are in place (e.g. due to variables not included in the model).

The algorithm pipeline consists of 5 distinct rules especially designed to provide one or several recommendations for each problem area, depending on percentage increase needed. For interventions to be considered as recommendations they will need to pass rule number 2. The processes running in rules 3 and 4 focus on ranking the available interventions according to impact, delay and cost. The last rule sets out that recommendations with higher impact, available resources, small lag/delay and lower cost will be prioritised against others of less importance.

The structure of the algorithm has been designed as follows:

1) The first rule ranks the interventions according to impact (how much it will increase returns). Interventions with higher impact should be considered first against others with lower impact. All interventions are classified by impact rank, the one with higher impact is classified as trial 1 intervention and will pass to the next stage. The second in the ranking is given trial 2 and so on;

2) In the second stage, trials are weighed against the availability of resources. When resources are still available trials are taken to the next stage. However, if resources are not available to implement this trial becomes blocked (i.e. if there aren’t enough resources for trial 1 then it becomes blocked 1). The same process happens with the other interventions in the ranking list (from stage 1);

 3) The next step includes assessing how much time would be required for trials to produce the desired effect (e.g. operational lag + response lag). This rule generates a second ranking system and provides information to the user about how much time it will be required to see the effect from the intervention;

 4) This stage of the process provides the information about the estimated cost for each trial that have passed rule 2 (available resources) and now have information about the lag appended. As for the previous rule, the 4th stage provides a ranking system that the user can use to base their decision on the cost;

5) The final stage concludes the selection process by producing a list of recommendations ordered by impact, lag and cost. Recommendations or combined recommendations with higher impact, available resources, small lag/delay and lower cost will be advised. Users can decide to implement these or choose other recommendations from the list.

Figure 6 – Diagram with intervention algorithm process flow with 5 steps (impact, resources, lag, cost and proposed).

# Visualisation Tool

During the live collection period, simulation data (i.e. response profiles) and other operational indicators will be used by the RCA as a basis for comparison against live return data, so that we can tell where we need to put in extra effort to meet our quality targets of: 94% overall response; at least 80% response in every local authority and reduced variability.

It is anticipated that the success of the first predominantly online Census and its likelihood to meet the quality targets will be highly dependent on the way users and stakeholders view and understand daily data describing all the different aspects of a national scale operation.

As a business decision support tool, the RCA has been designed to optimise and maximise the impact of its analytical features (prioritisation, intervention and forecast algorithms). When all these components are combined with the user-friendly python dashboard, users are provided with reliable information about census return and response patterns, different measures of statistical quality of the census data, and proposed actions to maximise response during live operation (Butterworth, 2019).



Figure 7 – User-researcher analysis into displaying interventions - “past, present and future” model

A specifically designed back-end and front-end functionality allows the RCA visualisation app to run smoothly and fully integrate information from different sources. Users are provided with the option to navigate between different levels of geography and to query data from the past as well as project their action into the future.



Figure 8 – View from the RCA Dashboard designed in python code showing operational MI (2019 Rehearsal synthetic data was used to create the visualisation).

All information provided by the RCA Dashboard has been designed to provide a user-friendly experience. Appropriate and statistically designed indicators provide constantly updated information required for a timely and well-informed decision-making process. The analytical features of the RCA Analytics (prioritisation, intervention and forecast algorithms) have been developed to provide information about different maximising response strategies. This information includes several business indicators (impact, delay, cost and availability) specifically designed to support users into their decision journey.



Figure 9 – View from the RCA Dashboard designed in python code showing recommendations to maximise response in Ceredigion 011G (2019 Rehearsal synthetic data was used to create the visualisation).

# Discussion

The 2021 Census will be predominantly online for the first time. As a result, we will have a lot more information much more quickly than ever before on who has submitted a Census questionnaire. Being able to use this information intelligently during live operations enables us to maximise the effectiveness and efficiency of our operation by tailoring interventions to the areas and populations where they are most needed.

This level of information has never previously been available, and the challenge in meeting response rate targets has never been so great. This tool plays a critical role in using the intelligence available during live operations to influence the volume and distribution of field officers conducting follow-up activities, the media and community engagement strategies and deployment of additional reminder letters and paper questionnaires. The 2021 Census is a very large logistical operation with up to 40,000 staff working in the field. Being able to target the resources effectively means that we can provide the best value for money while maximising the quality of the outputs we produce.

In its core, the especially designed algorithm analyses return rates, compares them to expected values calculated using a discrete event simulation, flags shortfalls and proposes interventions based on availability, lag, costs and impact. All analytical features and different components from the RCA Analytics (prioritisation, intervention and forecast algorithms) have been adjusted to provide users with up-to-date operational insight for each one of the maximising response strategies (RCA interventions: media campaigns; assisted digital; reminder letters, paper questionnaires and increased field staff hours) in a way that considers the relative impact from each strategy, as well anticipated operational constrains (business indicators: availability, delay/lag and cost).

A specifically designed back-end and front-end visualisation functionality allows the RCA to run smoothly and fully integrate information from different sources. The dashboard has been developed in a way that users are provided with the option to navigate between different levels of geography and to query data from the past as well as project their action into the future. All information provided by the RCA Dashboard has been optimised to deliver a user-friendly experience while providing all the necessary insight about the Census operation.

A combination between business and statistically designed indicators is expected to provide constantly updated information required for a timely and well-informed decision-making process. The RCA allows users to constantly monitor and maximise the available resources and consequently substantially reduce the risk of not achieving the census quality targets and the associated risk of overspending.

# References

Ardis, J., Shipsey, R., & Meirinhos, V. (2018) RCA tool discovery phase report (ONS internal paper available at <https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD_Stats_Design/Maximising_Response/RCA/Discovery/RCA_tool_Discovery_Phase_v1.0.docx>)

Barber, P. (2018). *Simulating the Census using a discrete event simulation* – *validation* (ONS internal paper available at <https://share.sp.ons.statistics.gov.uk/sites/MTH/Cen/2021/Sample_Design_and_Estimation/Field_Operations_Simulation/FOS%20-%20validation%20paper.docx>)

Bexley, S. (2018) *Assisted digital options – choices for assisted digital in 2021* (ONS internal paper – available on request)

Butterworth, D. (2019). *A task-focused design based on the user’s need to analyse the current, in order to plan for the future, based on findings from the past* (ONS internal document available at <https://collaborate2.ons.gov.uk/confluence/pages/viewpage.action?spaceKey=SDC&title=Census+RM+-+Design>)

Dini, E. (2018). *Hard to count index for the 2021 Census* (ONS internal paper 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))

Meirinhos, V. (2018) Maximising response – response profiles (ONS internal paper available at [https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD\_Stats\_Design/Maximising\_Response/Response\_Profiles/Response\_Profiles\_CRAG\_(0.2).docx](https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD_Stats_Design/Maximising_Response/Response_Profiles/Response_Profiles_CRAG_%280.2%29.docx))

Meirinhos, V. (2019a) Field prioritisation algorithm (ONS internal paper available at

<https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD_Stats_Design/Maximising_Response/FPA/Presentations/FPA_CRAG.docx>)

Meirinhos, V. (2019b) Maximising response – 2019 Rehearsal response profiles (ONS internal paper available at <https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD_Stats_Design/Maximising_Response/Response_Profiles/Response_Profiles_2019_CRAG.docx>)

Meyer, B., Mok, W., & Sullivan, J.(2015). Household surveys in crisis. *Journal of Economic Perspectives*, *29*(4), 199-226.

ONS (2019) Media rehearsal campaign (ONS internal paper – available on request)

Racinskij, V. (2015). Precision of population estimates for different levels and variability of response. Internal ONS paper

Salganik, M. (2017). *Bit by bit: social research in the digital age*. Princeton University Press.

# Appendix

## 8.1. RCA Prioritisation Algorithm - Rules and Processes

1. Read in all data: all returns and paper returns (this is data predicted by FOS which is our “target” data), files with sample size and geographies.

2. Leave only LSOAs that will be in Rehearsal.

3. Calculate online returns.

4. Calculate cumulative returns.

5. Calculate return rates.

6. Calculate online share.

7. Create live data by multiplying target data by a varying factor of 0.9-1.1 (cumulative returns)

8. Calculate live daily returns out of cumulative returns.

9. Calculate online live share

10. Calculate live return rates.

11. Calculate difference between observed and expected return rates.

12. Calculate final forecasted return rate and the difference between this rate and expected final rate. This difference will be used to prioritize LSOAs in the RCA output.

13. Add a variability within LA measure (this is a difference between a given LSOA and the top performing LSOA for that LA for a given day. LSOA flagged if variability greater than 10%).

14. Add a variability within HtC measure (this is a difference between a given LSOA and the top performing LSOA for that HtC for a given day. LSOA flagged if variability greater than 5%).

15. Aggregate LSOA level data to LA level.

16. Repeat steps 5, 6, 9, 10, 11, 12 for data aggregated LA level data.

17. Aggregate LSOA level data to National level.

18. Repeat steps 5, 6, 9, 10, 11, 12 for data aggregated National level data.

1. For the methodology behind the index, see the *Hard to Count index for the 2021 Census* paper by Ercilia Dini (2018). [↑](#footnote-ref-1)
2. For more information on the methodology, see Barber, 2018. [↑](#footnote-ref-2)
3. The 2016 test was a small-scale test conducted by the ONS where 11463 participants were contacted and asked, “If tomorrow was Census Day, how would you prefer to complete your form?". Using a universe of 1789 online respondents this test revealed that all age groups between the ranges of 18 to 74 had a preference to respond online (above 75% preference) and only the above 75 preferred to respond on paper (55%). [↑](#footnote-ref-3)