**Field Prioritisation Algorithm**

**Victor Meirinhos**

Contents

[Contents 1](#_Toc2692007)

[1. Introduction 1](#_Toc2692008)

[2. Reducing variability 2](#_Toc2692009)

[3. Variables 3](#_Toc2692010)

[4. Field prioritisation 3](#_Toc2692011)

[5. Analysis 4](#_Toc2692012)

[6. Proposed model 6](#_Toc2692013)

[7. Limitations 8](#_Toc2692014)

[8. References 8](#_Toc2692015)

# Introduction

The Response Chasing Algorithm (RCA) and the Field Prioritisation Algorithm (FPA) are two products developed by the Statistical Design Collection team aiming to support Census Operations during the collection period.

The RCA is a decision support tool composed of 4 main components operating at Lower Layer Super Output Area (LSOA) and above: 1) the RCA Prioritisation Algorithm; 2) the RCA Forecast Model; 3) Intervention Algorithm and 4) the RCA Dashboard. The FPA is an operational tool which uses return rates at Output Area level and automatically integrates with the Field Work Management Tool (FWMT).

The RCA has been developed to monitor daily census returns and responses, to provide insight about non-respondents and to propose maximising response recommendations (Meirinhos, 2019). During the Census collection period, it is anticipated that this tool will be used by the Census Operations, Census Field Operations and other business teams to plan and manage their operations. Overall, it is expected that this tool will contribute to achieve the desired Census Targets related to maximising response (e.g. 94% overall census response, 75% online response, at least 80% response across Local Authorities).

A different approach was used to develop the FPA. This tool has been specifically designed to reduce variability within each Local Authority (LA) by specifically targeting low response in Output Areas (OAs). This tool has been designed to be fully automated and operate twice a week by prioritising field visits within determined Census Coordinator Areas (CCA). This means that visits to some addresses will need to be prioritised over other addresses. This might happen because some areas present a critically low level of response and/or there is a risk of increased variability.

It is expected that during the 2019 Census Rehearsal and the 2021 Census the FPA will start its operation 16 days after Census Day and will stop with the last batch of addresses provided to field officers at the end of the collection period. For this to happen, the maximising response team (responsible for the development of the FPA algorithm) and TMobile (responsible for the development of the FWMT algorithm) are currently developing the integration between the FPA and the FWMT.

It has been proposed that the FPA will feed OA priority codes (0 – no visits; 1 – normal visits; 2 – high priority visits) to the FWMT. These codes and relevant information will then be used and integrated into the FWMT algorithm to schedule field officers’ journeys within the areas in which they have agreed to work, Census Collection Areas (CCAs).

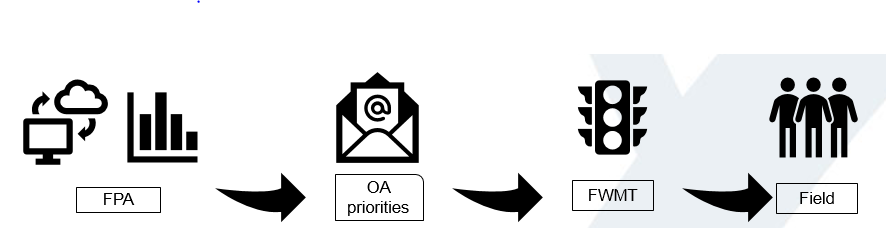


Figure 1. Field prioritisation algorithm workflow and integration with the FWMT.

# Reducing variability

One of the quality targets for the 2021 Census is to finish the collection period with a reduced variability across LAs in England and Wales. This requirement is mostly related to the estimation process that needs to happen after the Census Coverage Survey (CCS) and the fact that the precision of the estimation process after the Census is widely affected by the variance and level of response rate achieved at the end of the operation. The higher the response rate and the lower the variance the better the precision will be, assuming the same quality of the CCS across scenarios (Abbott and dos Santos, 2008; Racinskij, 2015).

From a maximising response perspective, it may seem tempting to seek a high response rate alone. However, the mean and variance of response rates are closely related and inter-constrained. Persistently increasing the response rate could eventually result in variance reduction. In fact, from a practical standpoint, hitting the census quality targets by only controlling the response rate would require a nearly perfect response rate. There is a known risk that even with a response rate as high as 97% the variance in responses could be large enough to cause the population estimates to miss the quality targets (Racinskij, 2017).

Therefore, similarly to the 2011 Census, a minimum of 80% at the LA level is proposed. Assuming that the variance at the low geography level is well-controlled and the responses between LAs are relatively uniform, overall response objective is proposed to be around 91 – 94%. In addition, and more importantly, for the 2021 Census, it is proposed to attempt to control variance while seeking the high response rate within the estimation strata or at the level close to it, say, broad age-sex group or even Hard to Count (HtC) group (ONS, 2016).

# Variables

The Statistical Design Collection team is currently developing two operational tools to maximise response and reduce variability during 2019 and 2021 Census operation. While the RCA has been developed to maximise response and reduce variability in higher geographies (LSOA and above), a new requirement has been identified by different users – maximise response and reduce variability in lower geographies (OAs).

To address this need, a new algorithm (FPA) has been proposed using live returns and the percentile distribution of OAs for each HtC classification. Synthetic data was developed to evaluate the consistency/appropriateness of the solution.

The synthetic data consisted of 199 OAs aggregated in 5 different CCAs within the same LA. Aiming to replicate the final period of the census collection process the data was formatted as cumulative OA return rates during a period of 10 consecutive days.

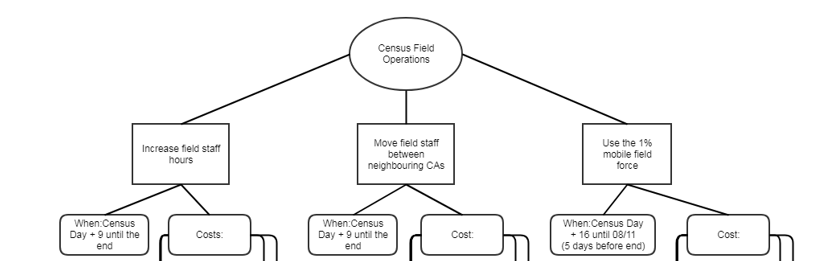
For the purpose of the analysis and consistency of the approach, the return rates from each OA were ranked according to the HtC classification from its respective LSOA. The percentages of OAs used in the split were approximately: 40% in HtC1, 40% in HtC2, 10% in HtC3, 8% in HtC4 and 2% in HtC5. This procedure aims to replicate the fact that areas more willing to respond to the census will have consistently higher return rates than areas with less willingness.

# Field prioritisation

To increase response while reducing variability within lower geographies, we propose that the FPA will be able to feed OA priority codes (0 – no visits; 1 – normal visits; 2 – high priority visits) to the FWMT. However, as field officers are expected to prioritise addresses within high level priority OAs (priority 2), this action cannot have a detrimental effect on the return rates from other areas/priorities.

Also, despite the fact that each coordinator will have the responsibility of reallocating resources within their area, we are confident that this mechanism will be vastly constrained and limited by the amount of increased hours that each coordinator will be able to negotiate with the field officers and/or the numbers of visits field officers will be able to perform daily. In areas where this option is not viable, alternative and much more costly solutions include field staff moves between CCAs or nationally (1% mobile force).

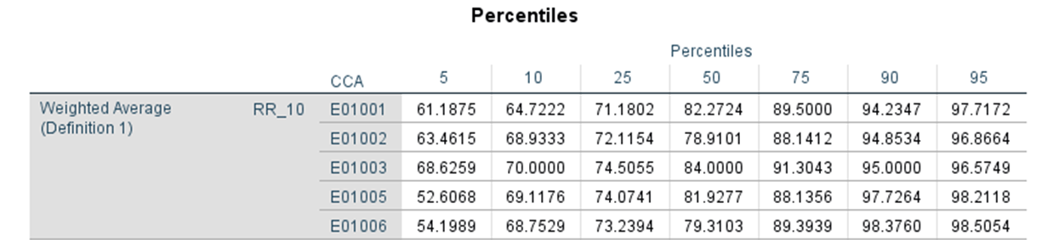
While all options are planned to be available during the 2019 and the 2021 operation, there is enough evidence suggesting that increasing field staff hours would be advisable when compared with other options. For example, compared with increasing field staff hours, moving field staff between neighboring coordinator areas will substantially increase the cost of the field operation and cause initial operational constraints due to the non-familiarity of the new field officer with the new census area.



**Figure 2**. Diagram displaying options and associated operational assumptions to increase field visits within each CCA (i.e. increasing field staff hours; moving field staff from neighbouring CAs, 1% mobile force).

# Analysis

The first analysis considered the CCA OAs percentiles to determine the sample of OAs in priority 2 (high), 1 (normal) and 0 (no-priority). By prioritising the percentiles 5 and 10 areas the objective was to reduce the range between top and low responding OAs within each CCA until that difference is lower than 10%. Likewise, percentiles 90 and 95 can be initially considered as clusters where response visits can be stopped/delayed and resources might be redirected to other areas.



**Priority 0**

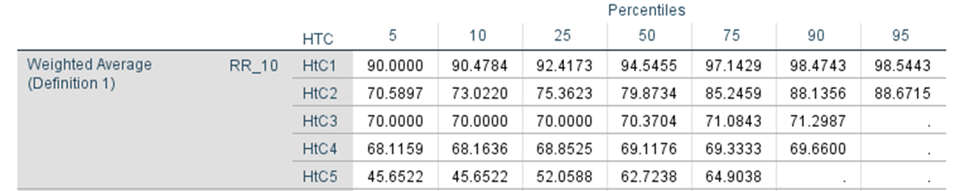
**Priority 1**

**Priority 2**

**Table 2**. Percentile distribution of return rates from 199 Output areas by CCA using the Weighted procedures. Priority 2 OAs can be selected from the percentiles 5 and 10; Priority 1 can be selected from the percentiles 25-50-75; and Priority 0 can be selected from the percentiles 90-95.

The second analysis considered a two-step process (i.e. within HtC variability correction/between HtC variability correction). In the first step, OAs from the HtC percentiles 5-10 where the return rate is 10% below the top responding OAs (in percentile 90 and 95) are flagged for priority 2. This procedure aims to identify HtC clusters where there is a need to reduce variability within the respective LA.

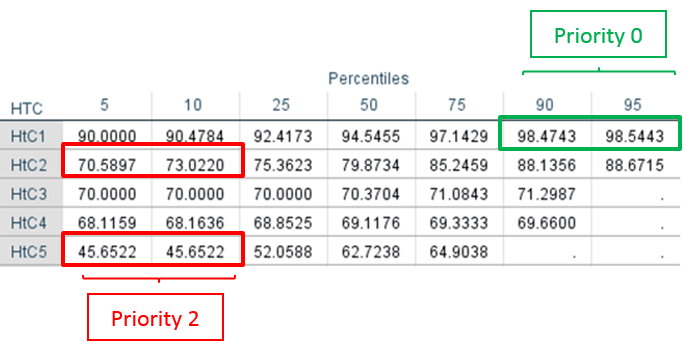
Priority 0



Priority 2

**Table 3**. Percentile distribution of return rates from 199 Output areas by HtC using the Weighted procedures. Priority 2 OAs can be selected from the percentiles 5 and 10 where the HtC interquartile range is higher than 10%.

In the second step, top responding OAs from HtC percentiles 90-95 if the HtC interquartile range is less than 10% are selected as a new cluster. These OAs will be given priority 0 (low) and will be the last addresses to be visited by field officers. This mechanism aims to reduce variability between HtC groups within a designated LA, provide extra-resources to increase the follow-up focus on priority 2 OAs while reducing the risk of affecting on-going maximising response strategies.



**Table 4**. Selection of Priority 2 and Priority 0 OAs using the two step prioritisation algorithm.

# Proposed model

The proposed model recommends that the FPA tool should target each CCA individually to more effectively reduce variance between and within the HtC classifications inside each Local Authority, in such a way that facilitates the estimation process that needs to happen after the Census Coverage Survey (CCS).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Reduce variance between to**  **< 10%** | **Reduce variance within to < 5 %** | | | | | | | |
| **HtC/Percentiles** | **5** | **10** | **25** | **50** | **75** | **90** | **95** |
| **1** | 1.05 | 1.10 | 1.25 | 1.50 | 1.75 | 1.90 | 1.95 |
| **2** | 2.05 | 2.10 | 2.25 | 2.50 | 2.75 | 2.90 | 2.95 |
| **3** | 3.05 | 3.10 | 3.25 | 3.50 | 3.75 | 3.90 | 3.95 |
| **4** | 4.05 | 4.10 | 4.25 | 4.50 | 4.75 | 4.90 | 4.95 |
| **5** | 5.05 | 5.10 | 5.25 | 5.50 | 5.75 | 5.90 | 5.95 |

**Table 5**. Displays all the relative percentile return profiles that could be used to classify OAs assigned to each CCA.

To achieve this, a balance between maximising response and variability control needs to be dynamically pursued during the short period where the FPA will be operational (from 16 days after Census Day until the beginning of the CCS).

The proposed model assumes the need to reduce the variance within each HtC classification inside each CCA to less than 5%, by continuously targeting the lower percentiles (5 and 10) as priority 2 (high) until the variance target within HtC is achieved.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Step 1 - Reduce variance within to < 5 %** | | | | | | | |
| **HtC/Percentiles** | **5** | **10** | **25** | **50** | **75** | **90** | **95** |
| **1** | P2 | P2 | P1 | P1 | P1 | P0 | P0 |
| **2** | P2 | P2 | P1 | P1 | P1 | P0 | P0 |
| **3** | P2 | P2 | P1 | P1 | P1 | P0 | P0 |
| **4** | P2 | P2 | P1 | P1 | P1 | P0 | P0 |
| **5** | P2 | P2 | P1 | P1 | P1 | P0 | P0 |

**Table 6**. Displays clusters for priority 0, 1 and 2 that could be used to prioritise OAs assigned in each CCA. Amber area shows the percentiles where priority 2 (high) can be found and Green area shows the percentiles where priority 0 (low) can be selected. This approach aims to reduce the variance within HtC groups to a level below 5%.

Once this has happened (the variance within HtC is below 5%) a new intervention starts to reduce the variance between HtC to values less than 10%. This approach also aims to increase response from the hardest to count areas and consequently increase the overall response rate. Combined, the two-stage approach aims to reduce the variance of Output Areas return rates within each Census Coordinator Area, by making sure that OAs from the same HtC classifications provide a similar response, and the ones below that desired level are effectively targeted with more visits.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Step 2 reduce variance between to < 10 %** | | | | | | | |
| **HtC/Percentiles** | **5** | **10** | **25** | **50** | **75** | **90** | **95** |
| **1** | P1 | P1 | P1 | P1 | P0 | P0 | P0 |
| **2** | P2 | P1 | P1 | P1 | P1 | P1 | P1 |
| **3** | P2 | P1 | P1 | P1 | P1 | P1 | P1 |
| **4** | P2 | P1 | P1 | P1 | P1 | P1 | P1 |
| **5** | P2 | P1 | P1 | P1 | P1 | P1 | P1 |

**Table 7.** Displays clusters for priority 0, 1 and 2 that could be used to prioritise OAs assigned in each CCA. Amber area shows the percentiles where priority 2 (high) can be found and Green area shows the percentiles where priority 0 (low) can be selected. This approach aims to reduce variance to a level below 10%.

# Limitations

The proposed model assumes that OA geography boundaries are required to match with the CCA boundaries. Current plans assume that this will only be possible for Hackney during the 2019 Rehearsal but will be extended to all areas in England and Wales in 2021. Further modelling needs to be conducted to develop a technical solution for areas where OA boundaries overlap with different CCAs.

Another limitation comes from the fact that the OA HtC classifications used in the present model are borrowed from the LSOA HtC where the OA is nested. This is due to the inexistence of a more appropriate proxy willingness measure for the lower geographies.

# References

Abbott, O. and Marques dos Santos, M.L. (2008) Precision of population estimates for different levels and patterns of census response. Internal ONS paper. Available on request.

ONS (2016). Census Transformation Programme | Initial user requirements for England and Wales: Response to consultation. Office for National Statistics, UK. Available from

<https://www.ons.gov.uk/file?uri=/census/censustransformationprogramme/consultations/assessmentofinitialuserrequirementsoncontentforenglandandwales/assessmentofinitialuserrequirementsoncontentforenglandandwalesresponsetoconsultation.pdf>

Meirinhos, V. (2019) Challenges in estimating the response to Census during the 2019 Rehearsal and the 2021 Census. Conference paper. RAS Conference. Newport. Available on request.

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

Racinskij, V. (2017). Census data collection field guidance. Internal ONS paper. Available on request.