Forecasting return rates in the 2021 Census

Brendan Davis, Orlaith Fraser

1. Background

During the 2019 Rehearsal the Response Chasing Algorithm (RCA) compared live and expected return rates to identify areas with shortfalls in response; to assign each a RAG status; to rank the shortfalls; and to propose interventions to mitigate them. The expected return rates were an output from the Field Operations Simulation (FOS) model, produced for each Lower level Super Output Area (LSOA). This provided a daily measure of progress against targets up to and including the day of reporting.

In addition, a simple forecasting measure was also used to estimate the future and final return rates at the end of the operational period. The forecasting measure derived the ratio of final expected return rate to current expected rate, and applied this ratio to scale the current live return rate to estimate a final rate. This method explicitly assumed perfect operational conditions from the next day forward; took no account of past performance as an indicator of future performance; and made no adjustment for deficiencies in the assumptions used in the modelling used to produce the expected return rates.

The RCA method and implementation used in the Rehearsal was successful in its core purpose to recommend interventions and report day-to-date progress. The RCA will be used in a similar fashion in 2021.

However, an improved forecasting method is needed to support the RCA – one that improves the accuracy of forward estimation of future performance; one that provides a view of progress that takes into account both the past and future operational capacity, and the differences in respondent behaviours compared to those assumed in the FOS modelling.

The proposed method will used an adapted version of the FOS, running at the end of every day, and incorporating emerging evidence from previous return patterns; the effectiveness of the deployed field resources to date; the projected field resources in the future; and an assessment of the behaviour of respondents relative to prior assumed behaviours - adjusting accordingly. In effect, this will mean reproducing daily future expected return values similar to those produced prior to the start of collection; learning from past performance and accounting for any deviation from assumptions by incorporating all new operational information as it emerges.

2. Discussion

2.1 The Field Operation Simulation Model
The FOS has been used to simulate the 2021 Census collection operation; inform the design of the wave of contact; and help to determine the estimated field and reminder resources required to meet the key census return rate targets. Additional outputs from the FOS include daily expected return rates at Lower level Super Output Area (LSOA) and the number of expected field visits at similarly local levels.

The FOS uses a set of statistical and operational inputs and implicit assumptions to model the expected behaviour of households and Census Officers during the collection period. These include, but are not limited to, the willingness of respondents to self-respond; propensity to respond / switch to respond via paper questionnaire; effectiveness of reminder letters in prompting a response; contact rate of field staff; and field visit durations.

2.2 FOS performance in the 2019 Rehearsal
Analysis of the FOS assumptions, as used for the 2019 Rehearsal, shows that broadly, for the assumptions for which the volume and quality of rehearsal data was sufficient to assess, the assumptions show good agreement with observed operational values. For example, reminder effectiveness shows good agreement across reminder and hard to count groups; field contact rates show broad agreement, with some local variations; and the profile of the propensity for paper response switching is as expected given the barriers to switching in conjunction with the voluntary nature of the rehearsal. A sample of assumption validation evidence is shown in Annex A.

2.3 Extending the FOS
Based on the 2019 performance, the FOS proves to be a robust and valid basis for adaptation and extension to the task of forecasting future return rates during the collection operation. With sufficient modification it should be possible to move beyond a single use model with one set of static outputs, to produce a streamlined model that adjusts dynamically as the facts on the ground change.

The choice of the FOS to determine forecast future returns also provides the means to produce dynamic sets of data to benchmark and monitor field performance - rather than a single static set that takes no account of practical constraints as they emerge, this approach will adjust to shortfalls in resource or variations over time or at a local level.

2.4 Approach
The assumptions used in the FOS (excluding field visit durations and field travel times) are stratified by Hard to Count (HtC) group. Two assumptions – self-response willingness and propensity for paper response – are further stratified within HtC by 10 age/sex categories. In order to preserve this second level of stratification during live operations it would be necessary to have timely and reliable response (not return) data in order to discount households in each age/sex category from the outstanding households that have yet to respond. The lag in availability of response data means that this is not recommended, and instead a simplified set of assumptions, stratified by HtC only, will be used. Prior to the final adoption of this approach, testing will be completed to ensure that this small element of simplification does not introduce significant deviation from the baseline level of response in any local area.

2.5 Method

The main adjustments to the inputs that will be considered are in the assumptions of respondent and field behaviour, and the level of field resource available at a local level.

Specifically, the following assessment and adjustment of assumptions are proposed:

i) Reminder effectiveness – no adjustments are proposed – the effectiveness assumed was proven to be robust in the Rehearsal; monitoring and adjustment during live operations is impractical because of the volume and phasing of reminders; and the complexity involved makes the differentiation of reminder and field interactions impractical.

ii) Propensity for paper response – based on rehearsal evidence, these assumptions appear to be robust and valid – it is anticipated that adjustments should not prove necessary. However, consideration will be given to supplementing HtC stratification with a degree of geographical stratification. This should capture and account for high concentrations of some demographic groups in which some extreme variation in propensity for paper response may be evident.

iii) Field contact rates – high likelihood for adjustment – it is likely that these will vary over time as the field force 'beds in'; there is likely to be significant variation from the rates observed in the Rehearsal (improved training and on boarding of staff etc) and from the underlying data from which the assumptions were derived; and there is a potential effect of contemporary differences in the frequency/time that people are at home should COVID related changes in home working patterns exhibit a long tail that persists deep into 2021.

Field contact rates will be assessed starting from the Tranche 2 field start date, but no adjustments will be made until at least the start of Tranche 3, and potentially not until one week after that. Subsequent frequency for adjusting rates is proposed to be at least weekly, and no more than twice weekly – more frequent adjustment, especially early in field operations may create significant volatility in day-to-day forecast changes.

iv) Willingness to self-respond – variation from assumed levels of self-response will be assessed from the start of operations, but no adjustments will be made prior to Census weekend. The phased delivery of initial contacts in the first 10 days, and the high probability of variation in levels of response in the first weeks mean that adjusted forecasts produced in the early self-response period will exhibit consequent day-to-day fluctuation and regional variability. Instead, analysis of self-response vs. the assumed levels will be conducted and reported as part of business intelligence measures, through governance channels. From the early days of operations, estimates of the likely number of reminders needed in each phase will be reported, with measures of confidence in the likelihood of staying within the capacity constraints of each phase. Adjustment forecast return rates, incorporating changes in willingness assumptions, will be produced from Census Day onwards. Self-response willingness will be monitored after Census Day by continuous analysis of response levels from all households that have yet to receive field visits or reminders

The availability of field resources compared to the planned capacity will be monitored as part of the field recruitment process. In the final phase of recruitment - as numbers should be approaching capacity - any shortfalls at a local level will be translated into an early 'live' field capacity measure, and used to model operations under reduced capacity scenarios. This will allow for early end-to-end operational forecasting from mid-February 2021, and once the initial contacts land, this early forecast will fold into full reporting of progress against expected values and full operational forecasting.

2.6 Contingent scenarios

In addition to the explicit monitoring and adjustment of assumptions; and the impact of field capacity variations, the use of the FOS during live operations also offers the potential to accommodate and adjust for the impact of other unplanned local factors effecting operations or respondent ability to respond – for example local restrictions on movement and interaction due to ongoing COVID related regulations. The means to accommodate and implement ad-hoc adjustments will be defined as part of the work to implement the adjustments already described above.

2.7 Outputs

Mid-February	early estimates of forecast vs. expected return rates for the full operational period reflecting reduced field capacity scenario (if applicable).
01 March (initial contacts land from 03 March)	next day+ return rate forecast incorporating field capacity variation from expected, based on latest recruitment status.
08 March	assessment of live self-response levels vs. expected; confidence of remaining within threshold capacity for reminder letters in each phase.
22 March (Census Day 21 March)	improved next day+ return rate forecast incorporating intelligence derived from analysis of self-response levels and ongoing fluctuations in field capacity.
23 March (T2 field follow-up starts)	improved expected field visit numbers, incorporating latest field resource levels.
24 March	early indications of likely variation in field performance from expected.
30 March/06 April (T3 field follow-up starts)	improved next day+ return rate forecast incorporating intelligence derived from analysis of field performance variation from expected.

3. Conclusion

This paper has outlined the proposed method to produce improved return rate forecasts and associated intelligence during live operations– using a modified Field Operation Simulation. This approach offers clear benefits compared to the approach used in the 2019 Rehearsal, and includes the potential to flexibly adapt to unplanned scenarios.

Members of the board are invited to review the proposed method, note the changes and improvements in the approach compared to the 2019 Rehearsal, and endorse the further development of processes and systems to implement the proposed method.
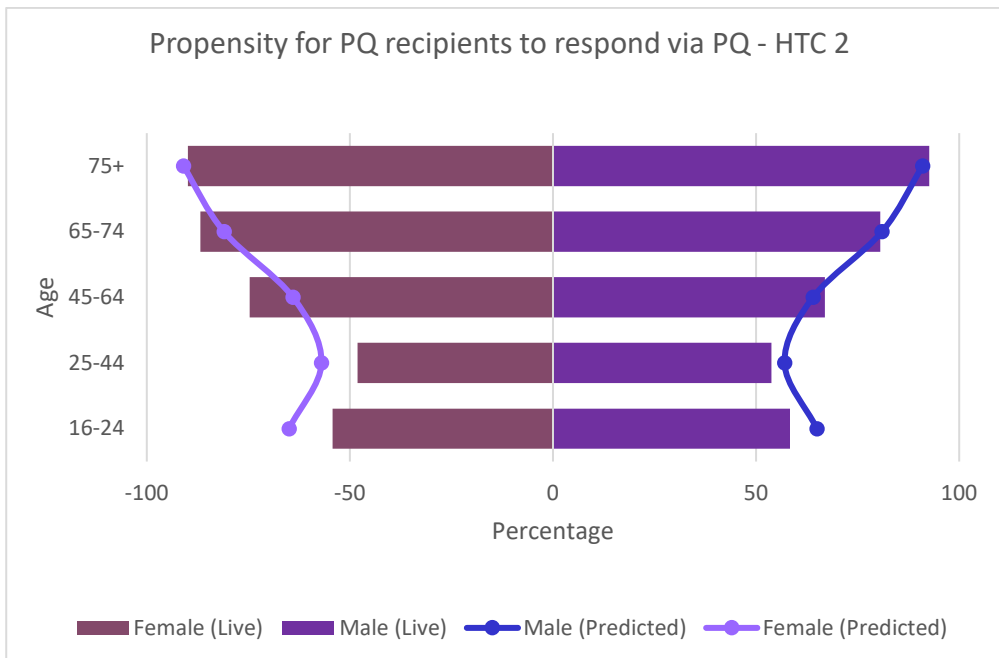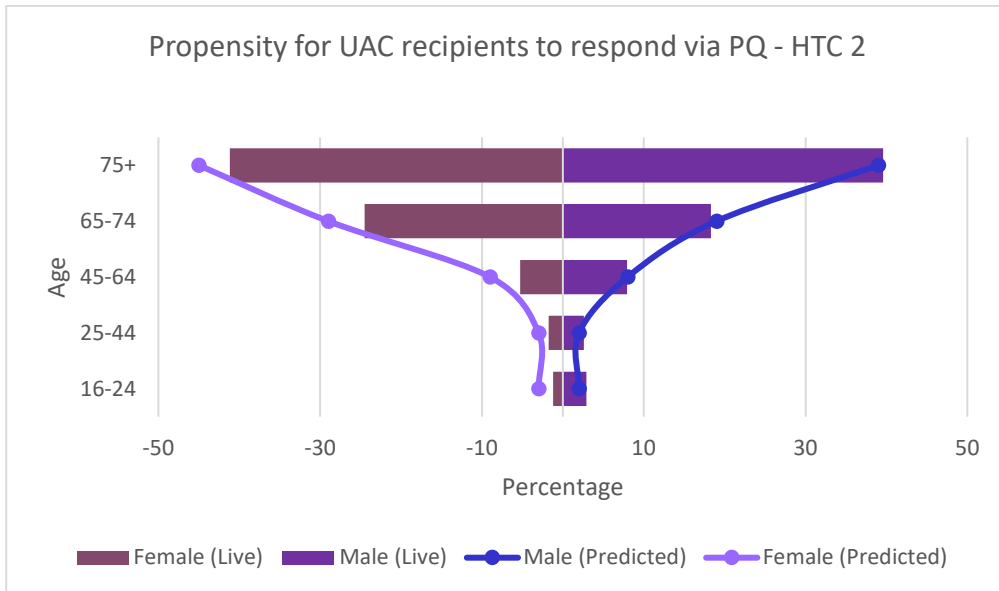
4. List of Annexes

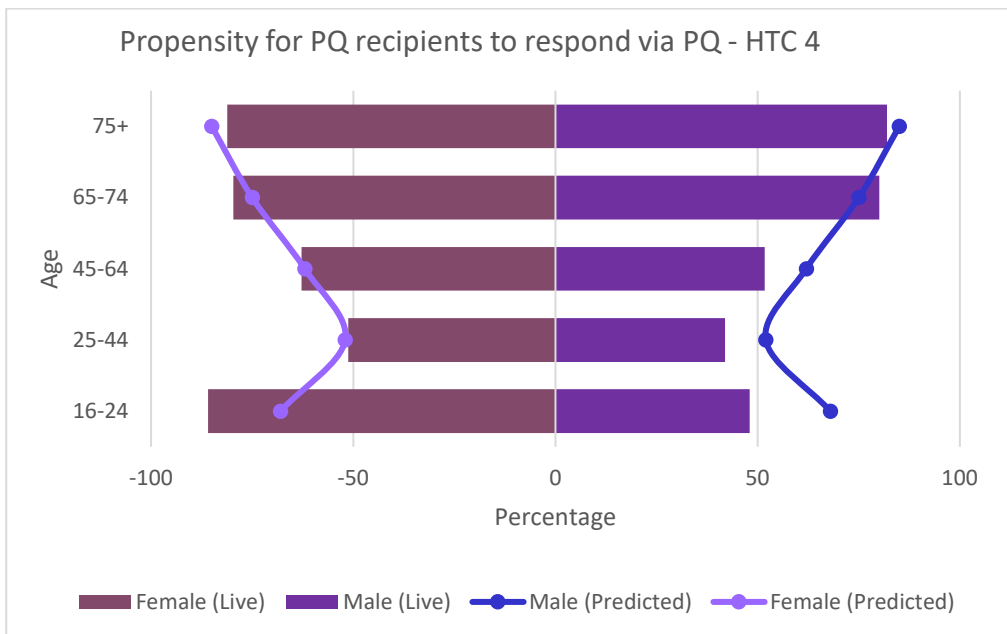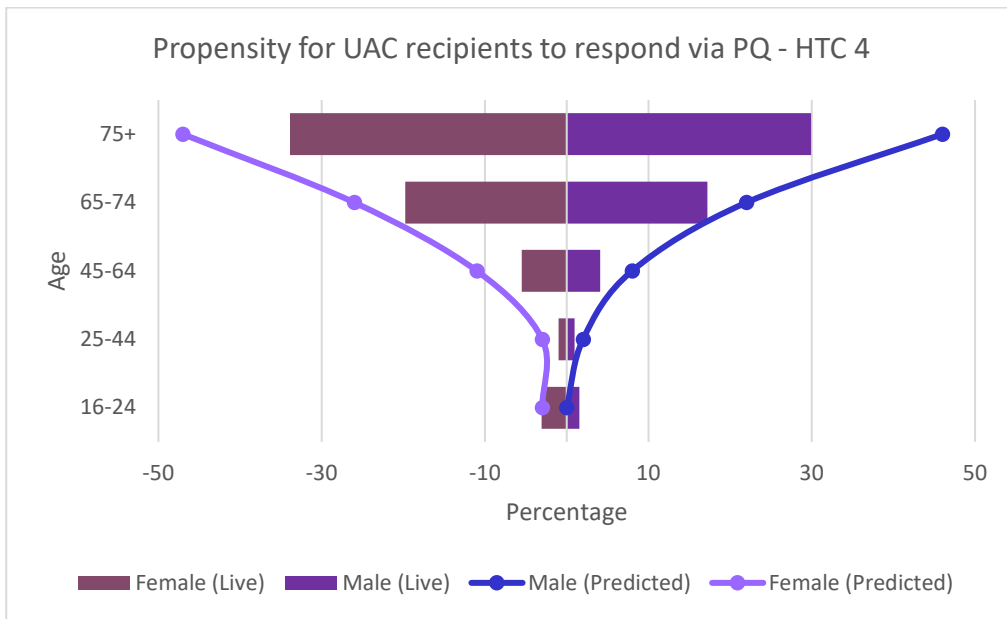Annex A: Validation of Field Operation Simulation assumptions

5. References

EAP117 Ward Kim, Barber Pete, Priestley Maria, Fraser Orlaith [2019], Simulating Census Operations to Inform Resourcing Decisions https://www.statisticsauthority.gov.uk/about-the-authority/committees/methodological-assurance-review-panel-census/

Annex A: Validation of Field Operation Simulation assumptions

As part of 2019 Rehearsal evaluation, where possible, some of the FOS assumption inputs were assessed and validated against respondent and field behaviours observed as part of rehearsal collection.

Propensity to response via Paper Questionnaire (PQ)

Propensity for UAC recipients to respond via PQ - HTC 2

Propensity for PQ recipients to respond via PQ - HTC 2

Propensity for UAC recipients to respond via PQ - HTC 4

Legend: Female (Live), Male (Live), Male (Predicted), Female (Predicted)

Chart: horizontal diverging bar chart by Age group (16-24, 25-44, 45-64, 65-74, 75+) with Percentage axis from -50 to 50.

Propensity for PQ recipients to respond via PQ - HTC 4

Legend: Female (Live), Male (Live), Male (Predicted), Female (Predicted)

Chart: horizontal diverging bar chart by Age group (16-24, 25-44, 45-64, 65-74, 75+) with Percentage axis from -100 to 100.

In Hard to Count 4, although the pattern of observed UAC to PQ switching broadly agrees, the assumptions overestimate the level of switching for upper age ranges. We can ascribe some of this difference to the voluntary nature of the rehearsal.

Effectiveness of reminder letters (non- field follow-up sample)

Live vs. assumed effectiveness of reminder letters
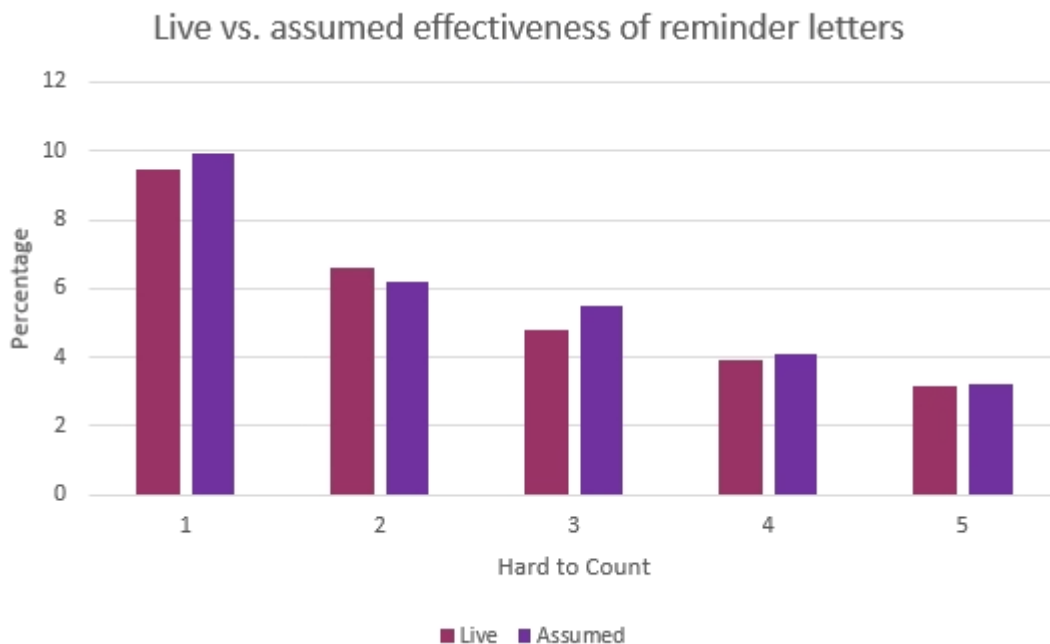
Effectiveness of paper questionnaire reminders (non- field follow-up sample)

Live vs. assumed effectiveness of PQ reminders
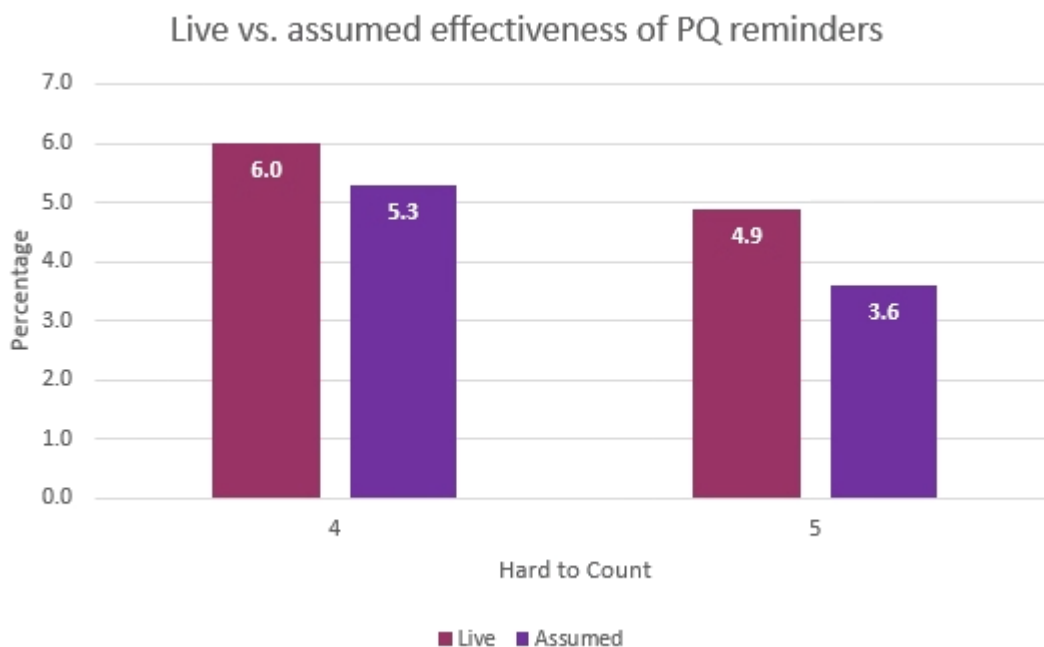
Field visit contact rates - live rates for London and non-London addresses compared to assumed rates