UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
UNECE Expert Meeting on Statistical Data Collection – Towards a New Normal?
26 to 28 October 2022, Rome, Italy

Session 3
19 September 2022

Achieving efficiency and representativeness through responsive design

Sabina Kastberg (ONS)
sabina.kastberg@ons.gov.uk

Abstract

At the UK's Office for National Statistics, it is essential we are as effective as possible when collecting representative survey data. To achieve this, we have implemented a responsive design within our telephone operations to increase efficiency on our Labour Force Survey.

A statistical model was developed based on information collected in the previous interview(s) and in the first week of data collection in the current field period. It produced the probabilities of how likely each case within each domain is to complete the interview in the second week of data collection and then prioritises cases based on their likelihood of responding, with the aim of achieving the same response rate with less interviewer resource.

Although the project aimed to increase efficiency, it was still important to ensure that non-response bias did not increase because of any design implemented. The project was designed to allocate cases to pre-defined mutually exclusive domains to ensure that non-response bias would not increase when case prioritisation was introduced.

Since this project, we have applied and extended this learning in research for how an adaptive survey design can be used on our transformed Labour Force Survey.

Both projects will be discussed during the presentation.

1. Introduction

Responsive design is "an adaptive data-collection approach that uses information available, both before and during data collection, to adjust the collection strategy for the remaining cases[1]." In most surveys all sample units are treated equally, and the same design features apply to all cases. Responsive designs assume that if sample units are treated in different ways, that resources may be targeted more effectively. The way in which the sample units are treated is defined before the survey starts but may also be updated via data that are observed during data collection.

The motivation to implement a responsive design came from the 2013 National Statistics Quality Review (NSQR) of the UK Labour Force Survey (LFS)[2]. The review recommended: "Explore the potential for adaptive fieldwork to focus resources more efficiently – in particular to: a) identify priority areas in advance of fieldwork; and b) explore opportunities for adaptive design during fieldwork." This paper focuses on the second of these recommendations where the aim was to determine whether data collection can achieve the same response rate more efficiently whilst also ensuring non-response bias is not negatively affected.

The LFS is the UK's largest household survey. Interviewing takes place over five consecutive quarters (waves) with the main mode of interviewing being face to face (CAPI) at wave 1 and telephone (CATI) for subsequent waves (waves 2-5). The capability to implement a responsive design given the current IT infrastructure at the Office For National Statistics (ONS) was greatest for the wave 2-5 cases provided the household had responded at least once previously. The functionality of the Blaise CATI call scheduler allowed the management of cases according to pre-determined rules.

Before the project was initiated, no demographic information about a household was used in prioritising cases. LFS cases were scheduled where every case started had an equal chance of being called. Prioritisation rules, based on a case's call history in that wave so far, were then used to determine priorities for subsequent calls. A program was run every five minutes to re-prioritise remaining cases.

The field period of the LFS is just over two weeks and each week there are three weeks of data collection proceeding simultaneously for different cases. At any given time there are:

- cases that haven't been tried (week 1 cases);
- cases that have already been tried for one week (week 2 cases); and
- cases that have already been tried for two weeks (week 3 cases).

To investigate responsive design, only week 2 cases were included in the scope for analysis. This means that only cases not yet contacted by week 2 of the field period were subject to the responsive design intervention. Any cases that have already completed an interview, already have an appointment scheduled for week 2, or have already refused were not in scope.

[1] **Responsive Collection Design Framework for Multi-Mode Surveys**, Francois Laflamme, Statistics Canada, April 2013
[2] http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-reviews/list-of-current-national-statistics-quality-reviews/nsqr-series--2--report-no--1/report---review-of-the-labour-force-survey.pdf

2. Methodology

The approach that was adopted was to calculate a response propensity for each case and then prioritise cases based on their likelihood of responding, with the aim of achieving the same response rate with less interviewer resource. However, although the project did not aim to reduce non-response bias it is still important to ensure that non-response bias did not increase because of any design implemented. The design therefore ensured that groups which have the potential for under-representation - either because they are under-represented at wave 1 of the LFS, or because they are more likely to drop out of the survey between waves – were not ignored. These groups needed to be prioritised equally with cases that were over-represented or less likely to attrite. Under-represented cases are likely to have a low probability of responding and would therefore be given a low priority within the call scheduler if the design was based purely on propensity modelling. This could have resulted in greater under-representation and cause further non-response bias in the sample than was present already.

To overcome this, the design allocated sample units to pre-defined mutually exclusive domains before any prioritisation was done. These domains were based on response rates of different groups, where the groups were put together based on household characteristics, and took into account representativeness at wave 1, as well as representativeness after attrition. The only information that was available on the non-responders at wave 1 was what could be found from the Census Non-Response Link Study (CNRLS)[3].

The domains were therefore based not only on response rates for different groups at waves 2+ (i.e. taking account of those most likely to attrite) but also considered factors which the CNRLS identified as a concern at wave 1 (i.e. taking account of household types which are less likely to take part at all). Additionally, domains were fixed groups that households could be allocated to, based on their household characteristics, rather than using any information gathered from survey questions that is likely to change from one wave to the next (employment status for example). All domains were weighted equally and propensities to respond were used to prioritise within the domains. So for example, a case which was the most likely to respond out of all cases in domain 1 was weighted equally against a case which was the most likely to respond out of all cases in domain 2, even if they had very different absolute likelihoods of responding. By following this approach, non-response bias should have not been increased by the introduction of the responsive design approach.

To establish a propensity model, call record and survey outputs data for the period between October 2015 and September 2017 was used. In total 30,403 cases from all waves of the LFS that were not contacted by an interviewer in the first week of the field period were considered in the analysis. Some of these cases were identified in several subsequent waves of the survey, therefore the data included repeated measures of the same household in different waves. Subsequently, a modelling approach to account for lack of independence between observations had to be used to model this data. One option is to use a Generalised Linear Mixed Model (GLMM) which can produce household-specific predictions based on fixed effects and random effects. However, due to computational challenges with implementing this approach at ONS it could not be considered. A computationally simpler approach was to use a Generalised Estimating Equations (GEE) approach which produces population-averaged predictions. In other words, households with the same covariate (independent variable) combinations have the same predicted values. In order to deal with the lack of independence between observations, the GEE approach specifies a working correlation matrix and has the advantage that the parameter estimates are consistent even if the working correlation matrix is mis-specified (Liang and Zeger,

[3] The CNRLS linked 2011 Census data to wave 1 LFS data to find out information about the non-responders. http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/non-response-weights-for-the-uk-labour-force-survey.pdf

1986). The dependent variable for this analysis was binary with a 1 indicating a full or partial interview had been achieved and a 0 otherwise (refusal or non-contact). A number of independent variables were considered for the analysis all of which were at the household level. However, some independent variables referred to the person characteristics of the Household Reference Person (HRP). A GEE approach with a logistic link function was fit to the data from the previous wave(s). Predicted probabilities for all households in the data were obtained. Records with missing data for age, sex or employment status variables were excluded from the analysis. The missing values are assumed to be Missing Completely At Random (MCAR). A total of 24,814 observations were included in the model (18.1% of total observations were excluded).

Table 2 shows all the variables included in the final model and their associated regression coefficient and p-values. The regression coefficients produced in the final model were plugged into the formula designed to calculate the response probabilities for 'live' cases in the implementation stages which is discussed in Section 3.

Table 2. Variables included in the model and their associated regression coefficients from the model

Variable	Coefficient	p-value
Intercept	-3.089	<0.001
Total calls this wave (4-6)	-0.305	<0.001
Total calls this wave (7-9)	-0.497	<0.001
Total calls this wave (10-12)	-0.276	<0.001
Total calls this wave (13-15)	0.788	<0.001
Total calls this wave (16+)	1.091	<0.001
Num. days case was called (2)	1.533	<0.001
Num. days case was called (3)	1.866	<0.001
Num. days case was called (4)	2.082	<0.001
Num. days case was called (5)	2.394	<0.001
Num. days case was called (6)	2.509	<0.001
Num. days case was called (7)	2.665	<0.001
Num. timeslots case was called (4)	-0.158	<0.001
Tenure (renting)	-0.579	<0.001
HRP's age group (35-44)	0.199	<0.001
HRP's age group (45-54)	0.458	<0.001
HRP's age group (55-64)	0.682	<0.001
HRP's age group (65-74)	1.087	<0.001
HRP's age group (75+)	1.302	<0.001
HRP lone parent (yes)	-0.275	<0.001
HRP's gender (female)	0.053	<0.001
HRP's work status (working)	0.106	<0.001
HRP's ethnicity E&W only (white)	-0.356	<0.001
HRP's ethnicity E&W only (ethnic minority)	-0.519	<0.001
HRP's ethnicity E&W only (missing)	0.270	<0.001

HRP's ethnicity S only (white)	-0.295	<0.001
HRP's ethnicity S only (ethnic minority)	-0.795	<0.001
HRP's ethnicity S only (missing)	0.0	<0.001

3. Implementation

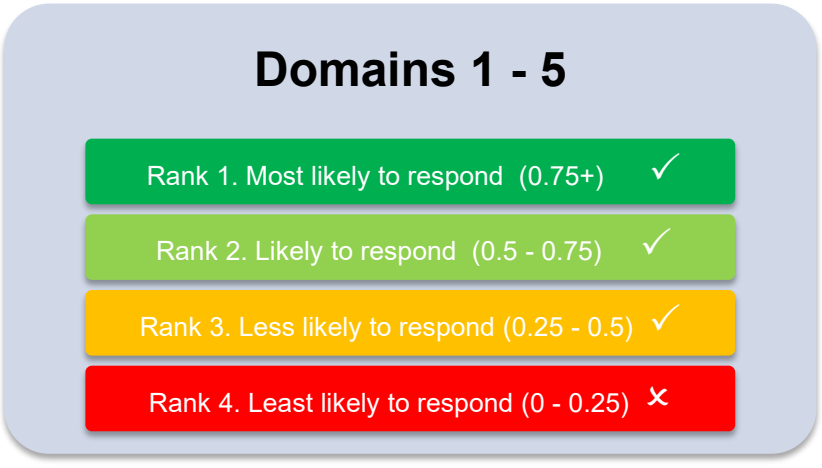
To implement the chosen responsive design, a number of processes have to be run ahead of the start of the second week of the field period. Once a given LFS week reached the Sunday of the first week a number of scripts in Blaise were run in the background in order to format the data and to carry out the required calculations. First the 'live' dataset from a given week was recoded in order to meet the requirements for calculating the predicted probabilities. The variables that were used in the probability calculation were dummy coded, i.e. split into multiple variables using a 1 and 0 coding convention.

Eligible cases (LFS W2+, that have no contact from an interviewer in the first week of the field period) were then selected, then the inverse logit function using the estimate values shown in table 2 was applied to calculate the probabilities.

Once household level predicted probabilities were calculated for all eligible households in the 'live' data, these predicted probabilities were then split into domains and then within each domain the predicted probabilities were split into four quartiles: cases with probabilities less than 0.25, 0.25-0.50, 0.50-0.75, and more than 0.75. The expectation was that around 25% of cases will be in the bottom 25th percentile from each domain in any given week.

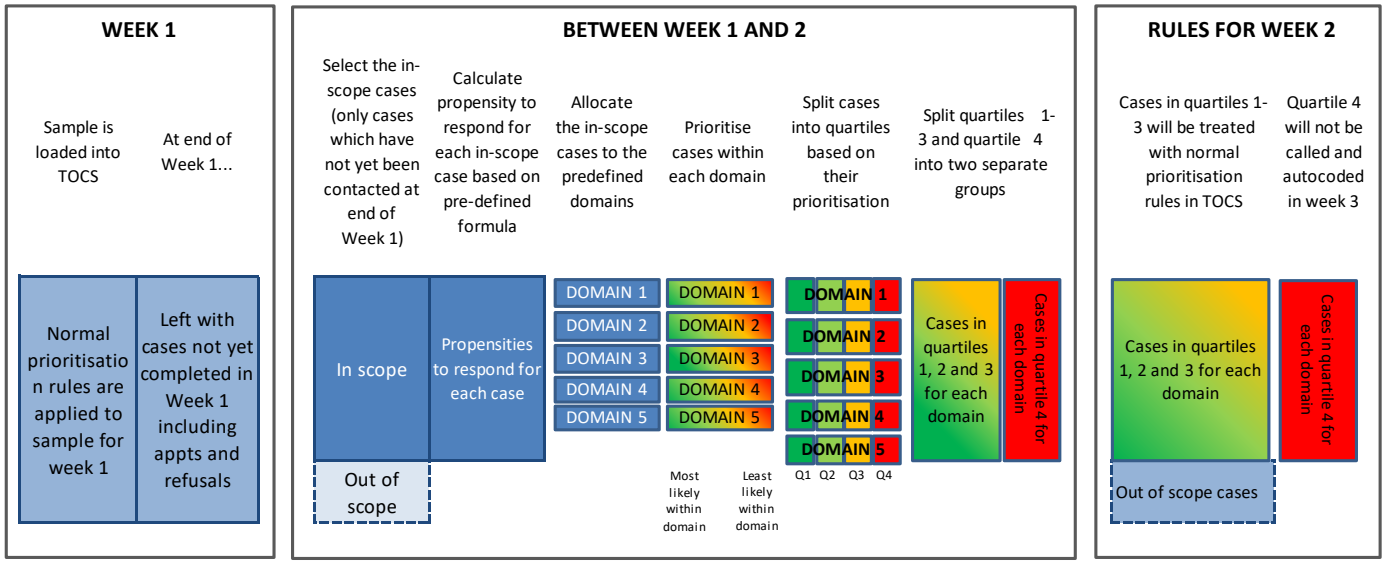
The unique probability thresholds for each quartile were then identified by applying the same methodology to each 'live' week's data. Following the identification of these unique thresholds, the definitions were then applied to the 'live' data and each case was assigned a rank from 1 to 4. Having to re-calculate the thresholds each week was necessary in order to achieve the 25% of cases in the rank 4 group (bottom 25th percentile) as weekly the data probability distribution was seen to be too volatile in order to use pre-set probability boundaries from the modelled data. The diagram below illustrates this approach.

Chart 1. Quartiles of the ranked cases within a given domain



Once the eligible cases from the live data were classified into Main (ranks 1-3) and OUT (rank 4) groups the cases that belonged to the OUT group were removed from the Telephone Operations Call Scheduler (TOCS) in week two and automatically coded out as final outcome 'non-contact' at the end of the field period, therefore they were automatically rotated to be tried again into the next wave. Cases in the Main group were treated exactly the same as they would have normally been treated in terms of prioritisation in relation to the new work, appointments etc. Therefore, those cases were called as normal following the regular TOCS scheduling rules along with the out of scope cases from the same reference week period. Chart 2 shows the implementation process visually.

Chart 2. Full operational implementation process of the responsive design



4. Evaluation

The responsive design started in Telephone Operations (TO) in August 2018 and has now been implemented as business as usual. The Trial evaluation discussed in this chapter will be looking at the data for the period between 20th of August 2018 and 15th of January 2019.

The success criteria for the intervention is the following:

(i) the final response rate for in-scope cases is higher than would have been expected for in-scope cases before the intervention

(ii) the response rate for in-scope cases is equal to the response rate that would have been expected for in-scope cases before the intervention and is achieved earlier on in the second week of data collection.

(iii) Calls to households are made more efficiently, i.e. initial contact is achieved quicker and fewer calls are made to households during the intervention period.

(iv) Cases across both the in scope and out of scope cases are processed more efficiently during the week, e.g. because less resource is needed to work on the in- scope cases because they are completed more efficiently after controlling for available staff numbers and other workloads.

(v) Additionally, the design aims not to negatively affect the non-response bias in the LFS. To evaluate whether this has happened, the characteristics of respondents in Waves 2-5 main LFS will be looked at for the time period where the design is implemented compared to the same period in the previous year. The proportion of households in each of the pre-defined domains will be used as a way of assessing whether the profile of respondents has changed because of the intervention.

Summary of the main findings:

- Response remained stable and this is a success – the primary aim of the project was to improve how efficient we are at collecting data, not to achieve higher response (response improvement was more of a secondary ambitious aim).
- Higher proportion of cases between Wednesday and Saturday (during their first week of the field period) are coded out as achieved interviews during the trial period, when compared to the previous year. This could be attributed to the responsive design as the old work is pushed out by TOCS between Wednesday and Saturday and as a consequence of better prioritisation of the old work in place more resource was attributed to achieve outcomes for the new work quicker. Results showed that in the second week the proportion of achieved interviews was almost equal in both trial period and the previous year, hence higher overall response rate on the LFS was not achieved.
- Significant decrease in the number of calls to first contact on LFS W2+ cases in TO.
- Significant decrease in the total number of calls made on LFS W2+ cases in TO.
- Significant increase in the total number of contact calls on LFS W2+ cases in TO.
- All the call result outcomes were significant after taking into account other ad-hoc surveys and staff hours worked.

Overall benefits:

- ✓ Creates resource availability to work on other surveys, without compromising LFS
- ✓ Mitigates the effect of staff shortages
- ✓ Fewer phone calls with a “no answer” for Week 2 work (which has boosted interviewer morale).