# Methods for reducing non-response bias due to the suspension of face-to-face interviewing

Bruce Fraser, Kristy Naylor and Alex Stuckey, Methodology Division, Australian Bureau of Statistics

## Summary

As a consequence of covid-19, the Australian Bureau of Statistics (ABS) has suspended face-to-face Computer Assisted Personal Interviewing (CAPI) for ABS household surveys. This suspension has caused response rates for new sample selections for the Labour Force Survey (LFS) and the Survey of Income and Housing (SIH) to decline by 12 - 15 percentage points. Without the CAPI option, respondents will report through either an online form or a Computer Assisted Telephone Interview (CATI).

These surveys are important for measuring the impacts that covid-19 has had on Australians. In order to maintain comparability with previous estimates, the aim is to remove any additional non-response bias caused by the suspension of face-to-face interviewing, but to the extent that there is residual non-response bias present in past estimates, we ideally want to retain this bias. This point is most relevant for new sample for the Labour Force Survey (LFS), where there is evidence of a small bias for respondents who are in the sample for their first month. Most of this bias from the incoming sample is ultimately removed by composite estimation before estimates are published.

The problem is complex because we expect that the probability of responding without the option of face-to-face interviews will depend on the survey variables of interest (specifically labour force status and household income), and additionally there has been a real-world change in the variables of interest (a decrease in employment and in household income) that coincides with the withdrawal of face-to-face interviewing.

There are three approaches proposed to address the additional non-response bias caused by the suspension of face-to-face interviewing. These are explained below and it is possible to use combinations of the approaches in a final solution. They are:

* matching administrative data to selected addresses and the population frame of addresses;
* matching respondents between cycles to observe the characteristics of selections that responded in one cycle but not another; and
* measuring the characteristics of face-to-face respondents in previous cycles of the survey.

As well as these approaches, there is also the default option of using the existing non-response methods. This is the preferred option if analysis concludes that the existing methods are effective at treating the additional non-response bias.

This is still work in progress although analysis has been performed for the April through July Labour Force Survey data with specific treatments recommended and implemented for LFS estimates.

## Description of proposed methods

#### Option 0: Use existing non-response methods

The ABS is confident that existing non-response methods are effective for the “business as usual” levels of non-response observed when respondents have a full choice of reporting modes (online, CATI and face-to-face/CAPI). When face-to-face interviewing is offered, the Labour Force Survey (LFS) achieves a response rate of approximately 92%, and the Survey of Income and Housing (SIH) achieves approximately 81%.

As a monthly survey, the LFS uses the same non-response treatment each month, which is calibration to age / sex / region population benchmarks. Region is state (8) split by capital city and rest of state for the six largest states, plus Northern Territory (NT) and Australian Capital Territory (ACT), giving 14 regions in total. The assumption underlying this method is that correcting for age / sex / region differences is sufficient to treat any non-response bias, or that non-response effectively happens at random within age / sex / region subpopulations. While this assumption seems reasonable for 8% non-response, it may not be reasonable if non-response increases to 20% due to the suspension of face-to-face interviewing.

The Survey of Income and Housing, along with other Special Social Surveys that repeat with a frequency of two or more years between cycles, are re-analysed each cycle to determine the best set of benchmarks to use in calibration. Age, sex and region are standard benchmarks, and additional benchmarks such as labour force status and the socio-economic ranking of the location of the household are also considered and often utilised. It is not uncommon to find that age, sex and region alone are not sufficient to treat non-response bias for response rates of 80% or lower, in particular residents of areas with low socio-economic status (or high levels of disadvantage) tend to be under-represented.

A calibration treatment, as part of the existing non-response adjustment toolkit, is an attractive option. In developing and trialling a calibration solution the following benchmarks will be considered, as previous knowledge suggests they are related to non-response. The following list is not exhaustive but summarises what are prima facie considered to be the most promising for explaining non-response: age / sex / location; labour force status; socio-economic status; highest educational attainment; country of birth; household size; household type.

One issue with trying to use a large number of benchmark variables to adjust for non-response using calibration is the limited number that can used without increasing variance and causing large changes in weights for some respondents. One approach that may help here is to perform an initial non-response adjustment using an estimated response propensity – for example using socio-economic status, education, country of birth and household type; and then a final calibration step, which for example may be based on age / sex / location and labour force status.

#### Option 1: Use of administrative data for calibration

A relatively straightforward extension of option 0 is to use administrative data, linked to the selected sample, for calibration. This allows for the use of data items that are strongly correlated to the survey variables of interest, for example payment of unemployment benefits for the Labour Force Survey and recipients of government payment and/or wage and salary earners for the Survey of Income and Housing.

The best correlation is achieved through data linkage at person-level, however for practical purposes we will assess the merits of this approach through address-level linkage. There are two key reasons for this. First of all, the sensitivity of using personal administrative data, and the risks of disclosure of personal information, are reduced through address-level linkage. Secondly, if person-level linkage is required then this is an additional step that must be done after survey data have been collected, and so it adds delay to finalising the weighting and publishing survey results. Address-level linkage can be done as soon as the sample has been selected.

While the use of administrative data seems very promising, the statistical correlation is reduced both through address-level, rather than person-level, linkage; and through the administrative data being somewhat out of date compared to the survey data at the time that weighting is done.

#### Option 2: Matching selections between cycles

This option is one that can be considered for panel surveys, such as the Labour Force Survey (LFS), where once a household has been selected they remain in the sample for a number of cycles (for eight consecutive monthly cycles in the case of LFS).

In the specific case of LFS, the withdrawal of face-to-face interviewing in March 2020 had very little impact on the continuing sample, i.e. the sample that had commenced the survey in February or earlier. This is largely explained by the ABS actively trying to move households to the online or CATI reporting modes after their initial month. In February 2020, face-to-face interviewing accounted for 40% of all responses from the incoming sample, but only 6% of responses from the continuing sample. With the suspension of face-to-face interviewing, the overall response rates for the continuing sample were maintained, indicating that the 6% who would normally have chosen face-to-face interviewing continued to report though an alternative mode. The incoming sample for each month since February has been impacted. The impact in March was minor, as face-to-face interviewing was only suspended when March data collection was in its final week. The April incoming sample had the greatest impact, with non-response increasing by approximately 22 percentage points. This indicates that for April, roughly one-half of the 40% that would normally choose face-to-face interviewing transitioned to other reporting modes, while the other half became non-respondents. The situation improved in May, with the incoming sample response rate about 13 percentage points below normal, indicating about ⅔ of respondents who would have chosen face-to--face have transitioned to other modes, with about ⅓ becoming non-respondents. In July, there are four cohorts of sample that are impacted, the incoming sample for each of April, May, June and July. The ABS has increased the size of the selected sample to maintain the total responding sample size, but this does not prevent non-response bias. The impact across the four cohorts joining the sample between April and July varies from between 8 and 16 percentage points of additional non-response.

May also saw a good proportion of the April incoming sample non-respondents convert into respondents for May. There were 675 persons reporting in May who were non-respondents from the incoming sample in April, which is equivalent to 14% of the responding April incoming sample.

Once the May data were available they could be used to produce a non-response adjustment for the April data. By matching the May respondents to the April sample, we could estimate the probability of a response in April, given a response in May and the May labour force status (as well as age, sex and location). Once these response propensities had been estimated, they were inverted to provide factors that were applied to the April initial weights (i.e. the inverse probability of selection weights) prior to the final calibration step. This adjustment did make a noticeable difference to estimates, and on the face of it removed a bias from estimates compared to the alternate method based on calibration alone.

Table 1 illustrates the differences in response for each labour force status. Table 1 shows the proportion of May respondents who did not respond in April, for the sample cohort that joined the survey in April. The labour force status for each column is the labour force status reported in May. So for persons responding as unemployed in May, 21.0% did not respond in April. If looking at the capital city sample only then 19.4% of the May unemployed sample dd not respond in April, and outside of the capital cities 23.9% of the May unemployed sample did not respond in April. As unemployed persons are more likely to non-respond, we conclude that unemployed people are over-represented in the April non-response, and so remedial action has been taken in estimation to address the risk of under-estimating the true extent of unemployment.

### Table 1: Rates of May respondents missing in April, for the sample cohort commencing in April, by May labour force status

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **FT employed** | **PT employed** | **Unemployed** | **NILF** | **All persons** |
| **Capital City** | 12.0% | 12.1% | 19.4% | 9.8% | 11.5% |
| **Rest of state** | 15.9% | 13.4% | 23.9% | 16.3% | 16.0% |
| **All areas** | 13.2% | 12.6% | 21.0% | 12.3% | 13.0% |

#### Option 3: Measure characteristics of face-to-face respondents in previous cycles

Where we don’t have the option of using either administrative data or response from other cycles, there is no option to directly measure the characteristics of non-respondents. Without a field workforce we cannot even distinguish between non-response and sample loss (e.g. vacant dwellings).

Instead we use an assumption that the additional non-response caused by the suspension of face-to-face interviewing is comprised of households that would have previously responded by face-to-face interviewing. It seems reasonable to assume that households that would have elected to respond by either CATI or online form prior to covid-19, will continue to respond when selected post-covid-19. From the Labour Force Survey it then appears about ½ (in April) to ⅔ (in May, June and July) of the households that previously would have elected to respond face-to-face now respond through another reporting mode, while ⅓ - ½ become non-response.

Can we use past cycles of survey data to establish the ways in which the face-to-face respondents are different, using this information to treat the non-response arising from the suspension of face-to-face interviewing?

Table 2 shows CAPI take-up rates for the incoming sample to the Labour Force Survey (LFS), prior to the suspension of face-to-face interviewing. Age, sex and labour force status are the characteristics of the household ARA, or Any Responsible Adult from within the household who provides the survey information on behalf of the household. This table suggests in particular that household composition may be an important factor to account for, in addition to the usual benchmarks for age, sex, location and labour force status. Broadly the tables also illustrate that CAPI take-up is highest among the subpopulations where non-response is typically highest. This is not surprising as CAPI is often used towards the end of enumeration to gather responses during the final follow-up.

Past cycles of survey data can be used to assess how effective post-stratification or response propensity methods are for treating non-response that is correlated to the CAPI response mode. For example, we can drop all CAPI responses from past survey data, apply non-response adjustment methods, study the change in estimates and assess whether systematic bias is present.

Post-stratification would be the usual weighting approach, and the challenge is to find the best set of post-stratification variables to best address bias. Another approach is to use past survey data to fit a model to explain mode choice. This modelling could be used to assign each responding household a nominal probability of responding through each available mode in circumstances where face-to-face interviewing is available. We can then look at the characteristics of households with a high propensity for electing to use face-to-face interviewing compared to the characteristics of other households in sample. If there is a difference in the survey variable of interest that is not sufficiently explained by age, sex, location and other potential post-stratification variables, we can then look at increasing the weights of these high propensity for face-to-face households to represent the additional non-response. This approach is still problematic though, as to treat non-ignorable non-response we would want to use the survey variables of interest in the propensity modelling. For example, we would like to use household income (or say equivalised household income) to help predict the propensity to respond via face-to-face interviewing. But when applying the model to estimate propensities for the impacted survey data, we need to use the post-covid values of income.

To illustrate, consider a household with relatively high income pre-covid that has had income substantially reduced due to covid-19. Are the choices this household makes to participate in the survey and to choose a mode of response driven by their pre-covid income level, or their post-covid income?

### Table 2: Proportion of incoming LFS sample respondents (households) by mode of response, by demographic group (cells with high CAPI uptake are highlighted)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Subgroup** | **CATI** | **CAPI** | **Online** |
| **All households / persons** | **32.2%** | **37.0%** | **30.9%** |
| Age / sex of ARA | M 15-24 | 27.5% | 58.9% | 13.6% |
| M 25-34 | 30.7% | 46.4% | 22.9% |
| M 35-44 | 30.4% | 38.8% | 30.8% |
| M 45-54 | 28.9% | 38.1% | 33.1% |
| M 55+ | 31.9% | 30.9% | 37.2% |
| F 15-24 | 29.7% | 51.1% | 19.2% |
| F 25-34 | 29.1% | 44.2% | 26.7% |
| F 35-44 | 31.0% | 38.9% | 30.1% |
| F 45-54 | 31.4% | 32.9% | 35.6% |
| F 55+ | 37.8% | 31.7% | 30.6% |
| Labour force status of ARA | Employed | 30.2% | 37.1% | 32.8% |
| Unemployed | 28.6% | 44.8% | 26.6% |
| Not in the labour force | 36.1% | 36.0% | 27.9% |
| Location | Capital city | 33.4% | 33.0% | 33.5% |
| Rest of state | 30.0% | 44.0% | 26.1% |
| Household type | Couple only | 30.3% | 33.3% | 36.3% |
| Couples with children | 29.8% | 35.6% | 34.6% |
| Single person | 38.0% | 36.8% | 25.2% |
| Single adult with children | 32.8% | 42.9% | 24.3% |
| Other | 30.5% | 43.9% | 25.7% |

## Conclusion

This paper has briefly summarised a number of options for treating additional non-response caused by the suspension of face-to-face interviewing in ABS household surveys. While work is still underway, initial investigations suggest that the loss of face-to-face interviewing exacerbates existing non-response problems, with the persons and household least likely to respond in normal circumstances also most impacted by the suspension of face-to-face interviewing. Additional estimation treatments have been applied to monthly labour force data to remove the resulting non-response bias. It is likely that similar treatments will also be applied to other ABS household surveys that have collected data since March 2020.