Understanding Society
Working Paper Series

No. 2021 – 03

March 2021

High frequency online data collection in an annual household panel study: some evidence on bias prevention and bias adjustment

Michaela Benzeval[1], Jonathan Burton[1], Thomas F. Crossley[2], Paul Fisher[1], Colin Gardiner[3], Annette Jäckle[1] and Jamie Moore[1]

This draft: December 2020

[1] Institute of Social and Economic Research, University of Essex,

[2] European University Institute, Florence

[3] Ipsos MORI Social Research Institute

Understanding Society

Economic and Social Research Council

Non-technical summary

Understanding Society is the UK's largest household panel survey. From April 2020, participants in *Understanding Society* were additionally invited to partake in a series of web surveys (the *Understanding Society* COVID-19 Study) designed to capture higher frequency information during the COVID-19 pandemic. These data allow researchers and policy makers to explore how the pandemic is impacting individuals and their families across the UK. A key feature of the *Understanding Society* COVID-19 Study is that it is designed to be "representative" of the UK population, in the sense of allowing analysts to make unbiased estimates of averages, prevalences and other statistics for that population.

In any survey, biases can arise if some segments of the population are not adequately covered by the survey design, or if particular segments of the population are less likely to respond to the survey. To prevent bias, the COVID-19 Study invites the full range of Understanding Society participants – both those who regularly use the internet and those who do not – to take part. Furthermore, periodically a subset of web non-respondents are invited to a telephone follow-up survey. Among web non-respondents, non-regular internet users are particularly targeted with this second mode.

To further adjust for any bias arising from nonresponse, a weighting strategy is implemented that takes advantage of the rich background information available from past annual interviews of the same individuals in the main *Understanding Society* survey. In this paper we examine the efficacy of these bias reduction and bias adjustment measures.

We find that both the telephone follow-ups and weighting help to reduce bias, but that inviting those who do not regularly use the internet to the web survey appears to be of little benefit. Overall we find that the *Understanding Society* COVID-19 Study is representative of the target population.

High frequency online data collection in an annual household panel study: some evidence on bias prevention and bias adjustment

Michaela Benzeval[1]
Jonathan Burton[1]
Thomas F. Crossley[2#]
Paul Fisher[1]
Colin Gardiner[3]
Annette Jäckle[1]
Jamie Moore[1*]

This draft: December 2020

[1] Institute of Social and Economic Research, University of Essex,

[2] European University Institute, Florence

[3] Ipsos MORI Social Research Institute

Abstract: *Understanding Society* is the UK's largest household panel survey. From April 2020, *Understanding Society* participants were invited to partake in a series of web surveys (the *Understanding Society* COVID-19 Study) designed to capture higher frequency information during the pandemic. To prevent bias, the COVID-19 Study invites the full range of *Understanding Society* participants – both those who regularly use the internet and those who do not – to take part and, furthermore, periodically invites a subset of non-respondents to a telephone follow-up. To adjust for bias, a weighting strategy is implemented that takes advantage of the rich background information available from past annual interviews. We examine the efficacy of these bias reduction and bias adjustment measures. We find that both the telephone follow-ups and weighting help to reduce bias, but that inviting those who do not regularly use the internet to the web survey appears to be of little benefit.

Keywords: web survey, mixed mode data collection, weighting, population inferences

Acknowledgements: The *Understanding Society* COVID-19 Study is funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). Fieldwork for the survey is carried out by Ipsos MORI and Kantar. *Understanding Society* is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research at the University of Essex. The research data are distributed by the UK Data Service.

---

*Author order is alphabetical.

[#] Corresponding Author: Department of Economics, European University Institute, Villa La Fonte, Via delle Fontanelle 18, 50014 San Domenico di Fiesole (FL), Italy. Thomas.Crossley@EUI.EU

1. Introduction

Understanding Society: The UK Household Longitudinal Study is a large, annual, multi-domain panel survey of people living in the UK. *Understanding Society* is based on probability samples and supports high quality population inferences.[1] It is widely used by decision-makers and researchers. Nevertheless, the rapidly changing circumstances of the COVID-19 pandemic created a demand for more rapid and more flexible data collection. In response, in March and April of 2020, the *Understanding Society* team developed and fielded the *Understanding Society* COVID-19 Study, which comprises a series of web surveys designed to capture higher frequency information from *Understanding Society* participants during the pandemic. At the time of writing, five such web surveys have been fielded, with sixth scheduled for November 2020.

Basing data collection during the pandemic on an existing household panel offers a number of important advantages. Data from the COVID-19 Study can be linked to prior information about respondents collected in the annual interviews of the *Understanding Society* Main Study dating back to 2009. These data provide important "baselines" and context. Similarly, future waves of the Main Study will provide an opportunity to study the long-run consequences of pandemic experiences. In addition, starting from a study that is based on probability samples, and which provides a wealth of background information with which to model response, provides a strong starting point for population inferences.

The validity of population inferences from *Understanding Society* have been extensively and positively assessed (Benzeval *et al.*, 2020). The COVID-19 Study departs from the Main Study in a number of important respects. The Main Study is normally a mixed-mode design, and it combines initial invitation to web and face-to-face interviews with follow up in alternative modes including telephone. The constraints of the pandemic mean that the COVID-19 study employs only web surveys,

[1] Understanding Society is built from four samples: the original British Household Panel Survey Sample, the General Population Sample, the Ethnic Minority Boost Sample and the Immigrant and Ethnic Minority Boost Sample. See Lynn (2009) for details.

with some telephone follow up. Moreover, the fieldwork period for the COVID-19 is very compressed, limiting the opportunities to obtain data from hard- to-reach and reluctant respondents. As a consequence, the COVID-19 Study is like to suffer greater nonresponse and attrition than the Main Study. Nevertheless, it is critically important to ensure that this new online data collection initiative is effective, efficient, and provides decision makers with high-quality data that supports reliable population inferences.

With regard to population inferences, the COVID-19 Study attempts to prevent bias through inviting the full range of *Understanding Society* sample participants (both those who are known to regularly use the internet and those who are not) to participate, and, at some waves, through a telephone follow-up of web non-respondents that reside in households where no one regularly uses the internet. Bias adjustment is undertaken through a COVID-19 weighting strategy that takes advantage of the rich background information available from the Main Study to model response to COVID-19 web surveys, conditional on participation in the Main Study (further details below).

The research questions we take up in this paper are:

- How effective were the bias prevention measures taken in the COVID-19 Study?
- How effective are the bias adjustments taken in the COVID-19 Study?

The first of these speaks to who to invite to a web survey, and to the value of the telephone follow-up. The second question provides a case study of the efficacy of bias adjustments for web surveys launched from a large annual panel survey, and of value of data derived from probability samples.

To answer these two questions, we focus on the first web survey in the COVID-19 Study and examine counter-factual samples. In particular, we compare the full sample, including telephone respondents, to the sample without telephone respondents and to a sample of only respondents who are regular internet users, defined as using the internet at least once or twice a week. The latter sample approximates the situation where data can only be collected from regular internet users, and it clarifies the value of attempting to recruit those who do not regularly use the internet to a web survey. Sets of Inverse probability (IP) weights are publicly available for the full sample, with and without telephone respondents. We created a further set of IP weights for the third sample, using the same methodology. This allows us to see the efficacy of bias adjustments with the different samples, and thus the interaction between bias prevention and bias adjustment strategies. To assess these samples and adjustments we compare estimates to benchmarks from Wave 9 of the Main Study.

To preview our results, we find evidence that in the COVID-19 Study, weighting and the telephone follow-up help to reduce bias, but inviting those who do not regularly use the web to the web survey appears to be of little benefit. We also find that weights based on basic demographics (as are often used with cross-sectional web surveys or non-probability samples) are not effective at eliminating bias.

Related Literature: Our analysis of bias prevention and bias adjustment measures in the COVID-19 Study relates to several literatures, including those on non-coverage and nonresponse bias in web surveys and on bias prevention and bias adjustment in surveys. Couper *et al.,* (2007) and Schonlau *et al.,* (2009) study non-coverage and nonresponse in a web survey launched from a established panel (the Health and Retirement Study) and the value of IP weights in adjusting for biases. They conclude that simply weighting on demographics is unlikely to provide sufficient adjustment. Schouten et al., (2016) examine whether bias prevention through choice and targeting of survey features, including mode, can reduce bias beyond what can achieved by ex post nonresponse adjustment alone. Such targeting of survey features might occur through adaptive or responsive survey design. They find evidence that more balanced response can lead to less bias, even after nonresponse adjustment. Most recently, Schaurer and Weiß (2020) study selection bias in online surveys on COVID-19-related behavior outcomes. In particular, they show that "big-five" personality traits are associated both with participation in a web survey and with compliance with risk-minimizing measures such as hand washing and minimizing social interactions.

Roadmap: The paper proceeds as follows. Section 2 describes the *Understanding Society* COVID-19 Survey in greater detail. Section 3 describes out our analysis strategy. Section 4 presents our results and Section 5 concludes.

2. The *Understanding Society* COVID-19 Study

2.1. *Understanding Society.*

Understanding Society: The UK Household Longitudinal Study is a large, multi-domain, longitudinal study of people living in the U.K. (University of Essex *et al.*, 2019). Study members are interviewed annually about health, education, employment, income, housing, family relationships, civic engagement and other topics. An attempt is made to interview all adult members of a household annually. *Understanding Society* began in 2009 but carries on from the earlier *British Household Panel Survey.* It employs a sequential mixed-mode design, with some respondents initially allocated to a web interview, and others initially allocated to a face-to-face interview with follow up in other modes. The *Understanding Society* sample was constructed from probability samples, and nonresponse and attrition have been carefully modelled.[2] A substantial body of research confirms the that the study continues to support valid population inferences (Benzeval *et al.*, 2020).

The *Understanding Society* team has undertaken a program of research and development on data collection modes, and this, in conjunction with the existing infrastructure for a mixed mode survey allowed *Understanding Society* to transition quickly to web and telephone interviewing in March 2020 and hence for fieldwork to continue through 2020, despite the COVID-19 pandemic (for further details, see Burton *et al.* (2020)). At the start of the pandemic Waves 10, 11 and 12 were in the field, while Wave 9 (2017/18) was the most recent wave of data that was publicly available. Nevertheless, *Understanding Society*, like other large longitudinal household studies, is not set up to field rapidly changing content in response to changing circumstances or provide data to researchers and policy analysis very rapidly. Consequently, in March 2020 the *Understanding Society* team took the decision to set up a new *Understanding Society* COVID-19

[2] The weighting strategy for Understanding Society is described in (Lynn and Kaminska, 2010).

Study, which would complement the annual interviews with more rapid web-based data collection.

2.2. The *Understanding Society* COVID-19 Study

Beginning in April 2020, *Understanding Society* sample members were invited to complete additional regular web surveys to track how the pandemic and associated policy responses were affecting them. The resulting data are the *Understanding Society* COVID-19 Study (University of Essex, 2020). The study is funded by the Economic and Social Research Council and the Health Foundation, with fieldwork undertaken by Ipsos MORI and Kantar.

The eligible sample was defined as all members of the *Understanding Society* Main Study who had not died, were aged sixteen or over, and had not emigrated as of in April 2020. All eligible individuals who belonged to "active" households, were invited to participate in the COVID-19 Study.[3] Pre-notification letters introducing the study were sent on 17 April. Respondents were offered a small financial incentive for each web survey.[4] Invitations to each web survey were then sent by email and/or SMS text message, or by post (depending on what contact details were available for each respondent). The design and implementation of the COVID-19 study built on research by members of the *Understanding Society* team on event-trigged data collection (see Jäckle *et al.*, 2019). Each web survey had a 7-day fieldwork period with reminders sent on days 2, 3, and 6.

Each web questionnaire was designed to take approximately 20 minutes to complete. The first web survey, which entered the field on 24 April, asked about coronavirus symptoms; the management of long-term health conditions; caring for others and loneliness; employment, earnings and household finances; food security; smoking, alcohol consumption and exercise; and

[3] Active households were those which had participated in one of the last two waves and had not explicitly withdrawn from the Study.

[4] At later waves they were offered the choice of taking the incentive or donating it to a NHS charity.

mental health. There was also a module on home schooling. For some topics (for example, employment) an attempt was made to measure a pre-pandemic "baseline" in January/February 2020 with retrospective questions. Data from the first web survey was released through the UK Data Service on the 29th of May.

Subsequent web surveys were fielded at the end of May, June, July and September, and further surveys are planned. These surveys maintained a "core" of questions to track change through the pandemic, but also include rotating content on, for example, housing, travel and transport use, religion, volunteering, and neighbourhood cohesion.

Further information on the *Understanding Society* COVID-19 Study can be found in the COVID-19 Study User Guide (Institute for Social and Economic Research, 2020).

2.3. Survey weights for the COVID-19 Study

Cross-sectional survey weights are released with each wave of the COVID-19 Study. These weights are derived via an adjustment to the cross-sectional weights provided with Wave 9 of the Main Study.[5] They map the cross-section of respondents to the relevant COVID-19 wave back to the target population at the time of Wave 9 (2017/18), updated for subsequent mortality and emigration, but not immigration.[6]

The weights provided with each wave of the COVID-19 Study are inverse probability (IP) weights. The probability of response to the COVID-19 wave in question is calculated as the product of the conditional probability of response to that COVID-19 wave, given response to Wave 9 of the main study, multiplied by the probability of response to Wave 9 of the Main Study.

[5] The Wave 9 weights are in turn adjustments to the initial design weights that take account of nonresponse and subsequent attrition (Lynn and Kaminska, 2010)

[6] Longitudinal weights that link the waves of the COVID-19 Study are under development.

The conditional probability of a response to a COVID-19 wave is modeled as a function of predictors drawn from the Main Study, and so observed for both COVID-19 respondents and COVID-19 nonrespondents. The availability of information on nonrespondents is one of the advantages of launching the COVID-19 study from a pre-existing panel. The potential set of predictors for response include basic demographics, household composition, economic variables and health variables. Further, both the survey statistics and econometrics literatures (Moffit *et al.*, 1999; Nicoletti and Peracchi, 2005) have suggested considering previous survey outcomes, survey design variables and survey paradata as predictors for response. These variables may be predictive of response, and may be correlated with outcomes of interest, but they are unlikely to be included in substantive social science or health science models. For example, whether a respondent has previously completed a web survey may be predictive for response to the current web survey and may be correlated with employment.

The conditional probability of a response to a COVID-19 wave is modeled as a Probit function. Final predictors are chosen from a large initial set via a Lasso procedure. Information on the most important predictors is presented below.

2.4. The COVID-19 Telephone Follow Up Survey

In an effort to capture respondents who would be unlikely to respond to a web survey, a telephone follow-up was conducted as part of the COVID-19 Study in May 2020.[7] This telephone survey ran contemporaneously with Wave 2 of the COVID-19 Study but was based on the Wave 1 (April) questionnaire.[8] It was issued to 3,432 Wave 1 (web) non-respondents, who, according to information collected in the Main Study, lived in a household where no-one was regular internet user. The invitation letter emphasized that non-internet users were an important part of society

[7] A second telephone follow up survey was fielded in November 2020.

[8] Adjustments were made to wording of the April Questionnaire to make it suitable for telephone, and interviewer instructions were added.

and that this was why they were being invited to a telephone survey, and they were offered a small unconditional incentive for completing the survey. Fieldwork began on 28 May and was completed on 7 June.

3. Analysis Design

To assess the bias prevention and bias adjustment steps taken for the COVID-19 Study, we focus on the Wave 1 web survey and the telephone follow up. We consider three potential samples. The first is respondents to the web survey who were regular internet users. This approximates the situation where data can only be collected from regular internet users. The second sample includes all respondents to the web survey, regardless of whether they were regular internet users. This is the standard sample that is released, along with appropriate IP survey weights. The third sample includes all responders to the web and telephone surveys. A weight has been released for this sample as well. For the first sample we created a set of IP weights following the same methodology as used for the other two samples. This allows us to see the efficacy of bias adjustments with the different samples, and thus the interaction between bias prevention and bias adjustment strategies.

To assess these samples and adjustments we compare estimates to benchmarks from Wave 9 of the Main Study. We follow the testing procedure described in Crossley *et al.*, (2020). Formally we test whether:

$$E\left[s_{t-1,i}R_{t-1,i}Y_{t-1,i} - s_{t,i}R_{t,i}Y_{t-1,i}\right] = 0,$$

where the response indicator $\mathrm{R_{t,i}} = 1$ if individual i responds to Wave 1 (of the COVID-19 Study) and 0 otherwise; and similarly, $\mathrm{R_{t-1,i}} = 1$ if the individual responds to Wave 9 of the Main Study. $\mathrm{Y_{t-1,i}}$ is an observation of any variable of interest, Y for individual i in Wave 9 of the Main Study. $\mathrm{X_{t-1,i}}$ is a set of predictors of response observed for both respondents and nonrespondents, prior to the realization of $\mathrm{R_{t,i}}$ (up to and including Wave 9 of the Main Study.) $\mathrm{X_{t-2,i}}$ is defined analogously. Note that these may contain lagged values of Y. $\mathrm{w_{t-1,i}}\left(\mathrm{X_{t-2,i}}\right) < \infty$ is the Wave 9 weight. This is the inverse of the Wave 9 response probability. $\mathrm{w_{t,i}}(\mathrm{X_{t-1,i}}) < \infty$ is the COVID-19 Wave 1 weight. The COVID-19 Wave 1 weight share is given by $\mathrm{s_{t,i}} = {}^{w_{t,i}}/_{\sum_i w_{t,i}}$ and similarly $\mathrm{s_{t-1,i}}$ is the Wave 9 weight share. The condition we test is implied by the joint null that:

$$E[R_{t-1,i}|Y_{t-1,i}, X_{t-2,i}] = E[R_{t-1,i}| X_{t-2,i}] = {}^1/_{w_{t-1,i}}, \text{ and}$$

$$E[R_{t,i}|Y_{t-1,i}, X_{t-1,i}] = E[R_{t,i}| X_{t-1,i}] = {}^1/_{w_{t,i}}.$$

That is, if response to Wave 9 of the Main Study, and to Wave 1 of the COVID-19 Study are both independent of $\mathrm{Y_{t-1,i}}$ given pre-response observables, then either combination of respondents (all Wave 9 respondents, or just those who also responded to Wave 1 of the COVID-19 Study) with the appropriate weights will provide a consistent estimate of $\mathrm{E}[\mathrm{Y_{t-1,i}}]$. The comparison of those two estimates provides a simple statistical test of the adequacy of the weights.

4. Results

Table 1 reports the prevalence of Wave 9 (UKHLS Main Study) information and response rates for Wave one of the COVID-19 Study and our three sample components. Column 1 indicates that 44,406 participants in the UKHLS Main Study were eligible for the COVID-19 Study.[9] Of these, 35,404 had Wave 9 (Main Study) information. The next three columns of Table 1, labeled (ii) through (iv) then describe our three sample components. Finally, Column (v) gives the descriptive statistics for the sample that pools all three components together.

[9] This number is correct at the time of writing. Over time, the number can drop slightly as information arrives indicating that a small number of these cases were ineligible (for example, because they were deceased by the time of the Covid-19 Study.) Weights are updated as this kind of information becomes available.

Table 1: Prevalence of Wave 9 (UKHLS Main Study) information and response rates for wave one of the COVID-19 Study and sample components. 'All eligible' is all eligible UKHLS subjects. 'Known regular internet users' are sample members reporting using the internet 1-2 times a week or more in UKHLS wave 9 (variable i_netpuse). 'Not-known regular internet users' are sample members reporting less frequent internet use, or for whom information is unavailable due to wave 9 item or unit non-response. 'Issued to Tel. Survey' are the COVID-19 Web Survey non-respondents who we subsequently attempted to interview by telephone (hence, they are also counted in one of the two earlier components). 'All' is respondents from all three components (i.e. the web and telephone surveys) combined.

	All Eligible	Respondents			
	(i)	(ii)	(iii)	(iv)	(v)
		Known Regular Net Users	Not Known-Regular Net Users	Issued to Tel. Survey	= (ii) + (iii) + (iv) All
N Eligible	44046	29740	14306	3411	
N Eligible with w9 info	35404	29740	5564	2955	
N Respondents	.	15514	2247	718	18479
Response rate	.	0.52	0.16	0.21	0.42
N Respondents with w9 info		15514	747	674	16935
Response rate, with w9 info		0.52	0.13	0.23	0.48

Column (ii) focuses on regular internet users. Again, these are those that used the internet 1-2 times per week or more. There were 29,740 regular internet users among our eligibles (67.5%). Since internet use status is determined from Wave 9 information, these all have Wave 9 information by definition. Among these, 15,514 responded to the first COVID-19 web survey, giving a response rate of 52%.

Column (iii) presents descriptive statics for those who are not known regular internet users. These 14,306 individuals include both those for whom Wave 9 information indicated they used the internet less than once per week (5,564), and those for whom Wave 9 information is missing (8,742). In total they account for 32.5% (14,306/44,046) of eligibles. However, of these "non-regular internet users", only 5,564 (39%) have Wave 9 information, so they are just 15.7% of eligibles with Wave 9 information. The overall response rate for this group was 16%, but among those with Wave 9 information the response rate it was just 13% (747/5564).

Column (iv) focuses on the 3,411 individuals issued to the telephone follow-up. Of these, 2,955 had Wave 9 information. The overall response rate to the phone survey was 21%, but for those with Wave 9 information it was 23% (674/2,955).

Finally, Column (v) considers the combination of these components. That is, we consider all eligibles, whether regular internet user or not, and define a respondent as one who responded to either the web survey or telephone follow-up (recall that all those issued to the telephone follow-up survey had previously been invited to the web survey, but not responded). The number of eligibles is therefore the same as in Column (i): 44,046, of whom 35,404 have Wave 9 information. The number of respondents is 18,479 (which is the sum across Columns (ii), (iii) and (iv): 15,514+2,247+718). Of these, 16,935 had Wave 9 information. With the telephone responses included the overall response rate is 42% and the response rate of those with Wave 9 information is 48%.

The key issue addressed in Table 2 is how "selected" the different potential sample components are, relative to the full set of eligibles. Descriptive statistics are for covariates measured at Wave 9 of the Main Study. This requires that we focus on the 80% (35,404 of 44,046) of eligibles with Wave 9 information (that is, those that responded to Wave 9 of the Main Study). The left-most column of Table 2 (labeled (i)) gives descriptive statistics for all eligibles with Wave 9 information. This is the most relevant group, as the Covid-19 Web Survey weighting strategy (described in Section 2) starts from the Wave 9 weight (and of course only Wave 9 respondents have a Wave 9 weight). Thus, given the current weighting strategy, population inferences from the COVID-19 Web Survey will be based on this group. Put another way, if we were able to obtain a response from all eligibles with Wave 9 information, no adjustment to the Wave 9 weights would be required.

The narrowest data collection strategy we consider is just issuing to regular internet users. Comparing column (ii) to column (i) shows that relative to eligibles, respondents drawn from the subset of regular internet users are more likely to be aged below seventy, more likely to have higher education, less likely to be Black Asian Minority Ethnic (BAME), and less likely to live in social housing. They also have somewhat higher incomes. Thus, perhaps unsurprisingly, a strategy of issuing only to regular internet users would result in a sample that is younger, more educated and more affluent than the eligibles. These imbalances would need to be overcome with adjustments to the Wave 9 weights.

Table 2: Socio-demographic characteristics for wave one of the COVID-19 Study. 'All eligible' (column (i)) is all eligible UKHLS Main Study members who were eligible for the COVID-19 Study. The other columns include respondents only. 'Known regular internet users' are web survey respondents who previously reported using the internet 1-2 times a week or more in UKHLS Wave 9 (variable i_netpuse). 'Not-known regular internet users' are web survey respondents reporting less frequent internet use, or for whom information is unavailable due to wave 9 item or unit non-response. 'Issued to Tel. Survey' are the COVID-19 Study web non-respondents who subsequently responded to the telephone follow-up survey. 'All' is respondents from all three components (i.e. the web and telephone surveys) combined.

	All Eligible	Respondents			
	(i)	(ii)	(iii)	(iv)	(v)
		Known Regular Net Users	Not Known-Regular Net Users	Issued to Tel. Survey	= (ii) + (iii) + (iv) All
Gender: Male	0.47	0.42	0.42	0.38	0.42
Age: 20-29	0.15	0.10	0.11	0.04	0.10
Age: 30-39	0.13	0.14	0.07	0.04	0.13
Age: 40-49	0.16	0.19	0.10	0.05	0.17
Age: 50-59	0.18	0.22	0.15	0.11	0.21
Age: 60-69	0.15	0.19	0.14	0.18	0.19
Age: 70-79	0.12	0.12	0.16	0.30	0.13
Age: 80-89	0.05	0.02	0.06	0.22	0.03
Age: 90+	0.01	0.00	0.01	0.06	0.00
Qualifications: Degree	0.39	0.51	0.22	0.18	0.48
Qualifications: A-level	0.22	0.21	0.18	0.14	0.21
Qualifications: GCSE or lower	0.39	0.28	0.60	0.67	0.31
Family type: Couple, kid(s)	0.24	0.27	0.17	0.06	0.25
Family type: Couple, no kid(s)	0.35	0.43	0.37	0.23	0.42
Family type: Single, kid(s)	0.03	0.03	0.02	0.03	0.03
Family type: Single, no kid(s)	0.37	0.27	0.44	0.69	0.30
BAME: Yes	0.20	0.12	0.17	0.16	0.13
Country: England	0.79	0.82	0.79	0.75	0.81
Country: Wales	0.06	0.06	0.06	0.08	0.06
Country: Scotland	0.08	0.09	0.09	0.09	0.09
Country Northern Ireland	0.06	0.04	0.06	0.08	0.04
Tenure: Owned	0.34	0.37	0.43	0.49	0.38
Tenure: Mortgage	0.39	0.44	0.38	0.11	0.42
Tenure: Rented	0.11	0.10	0.08	0.10	0.10
Tenure: Social Housing	0.16	0.09	0.11	0.29	0.10
Regular internet user: Yes	0.72	1.00	0.00	0.23	0.88
Household net income (£/month)	3556	3708	3687	1821	3635
Long-standing illness: Yes	0.35	0.33	0.49	0.55	0.34

Comparing Column (iii), the non-regular internet users, to Column (ii) reinforces this point.

Relative to respondents who are regular internet users, respondents who are non-regular internet users are older, have less education (for example, 60% have GCSE or lower, as opposed to 28% for the regular internet users), and more likely to be BAME. Thus, issuing to all eligible, not just those who are regular internet users, should lead to a somewhat more balanced sample.

Column (iv) shows that respondents to the telephone survey are even older than the non-regular internet users, as well as even more likely to have GCSE or lower education. They also have much lower household incomes and are much more likely to have a longstanding health problem. Adding the phone respondents should lead to further improvement in the sample, in terms of replicating the set of eligibles.

Column (v) then illustrates the combined effect of the two bias prevention strategies: a broad invitation (issuing to both regular internet users and to those who do not use the internet regularly) alongside mixed-mode data collection (following up some web non-respondents by telephone). The resulting sample is more similar to the set of eligible (in terms of observable covariates) than the sample in Column (ii) that results from narrower strategies for both invitation and data collection. Nevertheless, our final sample, in Column (v) is still significantly better educated than the set of eligibles (with 48% having a degree, versus 39%, for example). They are also less likely to be in social housing and less likely to be BAME. Consequently, the unadjusted Wave 9 weights (which map the eligibles to the target population) are unlikely to provide satisfactory basis for population inferences, even with this broadest sample.

To account for this, the COVID-19 Study has a bias-adjustment strategy based on adjusting the Wave 9 weights for nonresponse to the web surveys as well as for attrition between Wave 9 and the initiation of the COVID-19 Study. The following tables explore the use of that adjustment strategy in conjunction with our different samples. As our samples differ in the degree to which they attempt bias-prevention (through broadening the invitation, or employing telephone mop up), these tables show how bias adjustment and bias prevention interact.

As described in Section 2, the adjustments to the Wave 9 weights are based on predicted probabilities of response to the relevant web survey (or combination of web and telephone follow up). The predicted probabilities come from a Probit model, with covariates taken from information collected at Wave 9 of the Main Study. The covariates to be included are chosen by LASSO. The LASSO procedure selects a large number of Wave 9 covariates for inclusion as predictors. Table 3 shows the estimated (partial) effect on response probability of some of these covariates for our second sample (that is, both regular internet users and non-regular internet users, but considering that web response only). The estimated effects on response probability are presented as average marginal effects. For continuous predictors this is the average estimated derivative of the response probability with respect to the predictor, holding other predictors constant; for dummy variable (0/1) predictors this is the average change in the estimated response probability as the predictor changes from 0 to 1, holding all other predictors constant. Table 3 lists the twenty predictors with largest marginal effect sizes in absolute value.

Table 3: Top 20 Lasso response propensity model predictor variable marginal effect sizes. Sample of all web respondents (both known-internet-users and non-known-internet-users, but considering web responses only, not the telephone follow-up). All predictors in the model are coded as indicator variables. Hence, in each case the reference category is those sample members not identified in the variable name: for example, in the first instance, females. Variables are also grouped by subject matter.

Predictor variable	Marginal Effect	t-stat
Gender: Male	-0.07	-8.59***
Ethnicity: Irish	-0.09	-11.37***
Region: Northern Ireland	-0.07	-2.31*
Age band: 16-29	-0.11	-8.85***
Age band: 30-39	-0.06	-6.47***
Age band: 80+	0.10	0.03
Qualifications: GCSE or lower	-0.08	-9.28***
Occupation: Professional	0.09	0.04
Occupation: Administrative and secretarial	0.09	0.04
Occupation: Associate professional and technical	0.08	0.05
Standardised income decile: 6	0.06	5.30***
Standardised income decile: 5	0.06	5.01***
Standardised income decile: 9	0.06	4.14***
Reported income from savings and investment: Yes	0.08	10.39***
Tenure: Local authority rent	-0.05	-6.06***
HH type: 3 or more adults, no kids, incl. at least one couple	-0.06	-5.31***
Mode at wave 9: Web	0.23	32.08***
Email known at start of COVID survey	0.28	25.80***
Internet use: Less than once a month	-0.13	-10.25***
Internet use: Once / several times a month	-0.09	-11.76***

The two largest effect sizes in Table 3 have to do with the respondents' previous interactions with *Understanding Society*. In particular, those who had answered Wave 9 of the Main Study in web mode were, on average, 23% more likely to respond to the first web survey in the COVID-19 Study, all else equal. Second, eligibles for whom the *Understanding Society* team held an email address (and so were invited to the COVID-19 web surveys by an email containing a web link) were 28% more likely to respond to the first web survey in the COVID-19 Study, all else equal.

The aim of weighting adjustments is to eliminate bias. However, unequal weights are generally inefficient, so that the reduction in bias may come with a loss of precision (Denver and Valliant, 2018; Solon *et al.*, 2015). One advantage of a bias prevention strategy may be that bias can be eliminated with less loss of precision. Table 4 explores this issue. We consider the IP weights associated with three possible samples associated with the first web and telephone surveys in the COVID-19 Study. These are: just the regular-internet-user web respondents; all web respondents; and web and telephone respondents combined. In each case the associated weights are created in the same way, adjusting the Wave 9 weights with inverse probabilities of response to the COVID-19 survey, estimated as described above. We will subsequently assess the ability of weighting adjustments created in this way to eliminate biases in each sample. Here we consider the variability of the weights as a measure of the *potential* for precision loss. We present two measures. The first is simply the coefficient of variation (CV) of the weights. The second is Kish's DEFF which is a measure of actual variance inflation in a special case, and more generally provides some guide to how concerned we should be with the inefficiency.[10] For each sample we present two cases. The first is the "raw" weights, which are simply the Wave 9 weight adjusted by the inverse probability of COVID-19 survey response. The second is the "trimmed" weights, in which we bound the adjusted weights at 26 times the median adjusted weight. The trimmed weights correspond to what is actually publicly released by *Understanding Society* for the COVID-19 Study. Experience suggests that this bounding improves precision at the expense of allowing some bias. A potential benefit of bias prevention (though a broad invitation strategy or mixed mode data collection) could be that less trimming is necessary.

Table 4 shows that adding the non-regular internet users does not reduce the variability of the weights (in fact, by either measure the variability increases). Thus, a broader invitation strategy is unlikely to bring precision gains (for a fixed sample size). However, incorporating the telephone

[10] Kish's DEFF is the variance inflation factor for a sample mean from a stratified random sample with equal variances across samples. See (Denver and Valliant, 2018).

sample does substantially reduce the variability of the weights and particularly so for the raw weights, so that less trimming is required. This suggests that incorporating the telephone respondents may both improve precision and reduce bias. The latter effect would be in part because we need to do less trimming.

Table 4: Kish's DEFF estimating variance inflation due to use of the created inverse probability (IP) weights and Coefficients of Variation of such weights for each the cumulative COVID survey (respondent) dataset components. "Raw" reverse to the response probability from the response probability model. Trimmed is after trimming at 26 times the median response probability, which we do to control the variability of the weights.

	(i) Regular internet users		(ii) All web = (i) + non-regular internet users		(iii) All web + telephone = (ii) + telephone	
	Raw	Trimmed	Raw	Trimmed	Raw	Trimmed
DEFF	6.2	2.6	8.2	2.7	2.3	2.2
CV	249.5	140.1	306.6	156.7	138.0	133.6

Table 5 examines bias reduction through weighting adjustment directly. As described in Section 3, we consider the ability of each potential sample from the first survey (regular internet users only; all web responders; web plus telephone responders) to replicate Wave 9 population estimates for a range of covariates. That is, we use the Wave 9 survey reports for the relevant COVID-19 sample, plus the weights associated with that sample, and compare the resulting mean estimate to that obtained using the full Wave 9 sample (and associated weight). The validity of population estimates from the main study has been thoroughly investigated (Benzeval *et al.*, 2020) so the full Wave 9 estimates provide a suitable benchmark. The first column of Table 5 gives the full Wave 9 estimate. Subsequent columns give the biases (the difference between the benchmark estimate and relevant COVID-19 sample estimate). All of the covariates considered are binary, so that the weighted means are estimated prevalences, and the reported bias can be multiplied by 100 to give a percentage point difference.

We also test a second estimate from each COVID-19 sample. This estimate uses simple calibration weights instead of the refined IP weights developed for the COVID-19 Study. These match the relevant sample to weighted Wave 9 cell frequencies with cells defined by age, gender and education. The reason to consider the performance of these weights is that a number of web surveys conducted during the pandemic either use quota samples with quotas defined by these variables (Belot *et al.*, 2020) or use non-probability samples combined with calibration weights which match these variables to a benchmark such as the Labour Force Survey (Adams-Prassl *et al.*, 2020). Thus, the relative performance of such weights, with different samples, is of interest.

We divide the Wave 9 variables we consider into two sets. The first are those that are included in the response probability models. The second are variables that are not included in the response probability models. The latter stand in for survey target variables.

Table 5: Statistical tests of COVID-19 Study weight performance. We test whether differences between wave 9 weighted survey variable means / prevalences ('Wave 9: wt. est.) and similar estimates computed using COVID survey with calibration and inverse probability (IP) weights (respectively, 'C. wt. diff.' and 'IP wt. diff.') equal zero. We consider each cumulative COVID survey (respondent) dataset (Regular internet users, Web = regular internet users + non-regular internet users, Web and telephone), and both survey variables in the IP weighting model and variables in neither weighting model. `Core benefits' include Income Support, Job Seeker's Allowance and Universal Credit.

	Wave 9	Regular internet users		W1 web		W1 web and telephone	
Variable	wt. est.	C. wt. diff.	IP wt. diff.	C. wt. diff.	IP wt. diff.	C. wt. diff.	IP wt. diff
<u>In IP weighting model only:</u>							
Subject financial situation (SFS): living comfortably or doing alright	0.71	-0.05***	0.00	-0.04***	0.00	-0.04***	0.00
	(0.00)						
SFS: just about getting by	0.21	0.03***	0.00	0.03***	0.00	0.03***	-0.00
	(0.00)						
SFS: finding it quite/very difficult	0.07	0.01***	-0.00	0.01***	-0.01	0.01***	-0.00
	(0.00)						
Tenure: Owned	0.34	-0.06***	0.01*	-0.06***	0.01	-0.05***	-0.00
	(0.00)						
Tenure: Mortgage	0.34	-0.07***	-0.02***	-0.06***	-0.01**	-0.06***	0.01
	(0.00)						
Tenure: Rented	0.13	0.03***	-0.01	0.03***	0.00	0.03***	0.00
	(0.00)						
Tenure: Social Housing	0.19	0.09***	0.01	0.09***	0.00	0.08***	-0.01
	(0.00)						
Low skill occupation	0.38	0.05***	-0.01	0.05***	-0.01	0.05***	-0.01
	(0.01)						
Any savings income	0.36	-0.08***	-0.00	-0.08***	-0.01	-0.07***	0.00
	(0.00)						
Behind with some or all bills	0.06	0.02***	0.00	0.02***	0.00	0.02***	0.00
	(0.00)						
<u>In neither weighting model:</u>							
Income poverty	0.15	0.03***	0.01	0.03***	0.01	0.02***	-0.01
	(0.00)						
Receives core benefit	0.05	0.02***	-0.00	0.02***	-0.00	0.02***	-0.00
	(0.00)						
Behind with housing	0.09	0.03***	0.00	0.03***	0.00	0.02***	-0.00
	(0.00)						
Smoker	0.15	0.05***	0.02*	0.05***	0.02**	0.05***	0.01
	(0.00)						
Long-standing illness	0.37	0.04***	0.02**	0.04***	0.01	0.03***	-0.01
	(0.00)						

The first result to note in Table 5 is that the calibration weights do not work well. Even with bias prevention though a wide invitation strategy and multiple data collection modes (the full web plus telephone sample), there are significant biases in most covariates. For example, in the fraction of adults living in homes owned outright, the bias is 6 percentage points on a baseline of 34 percent. Biases in the incidences of capital income, smoking, and social housing are of similar magnitude, and all are statistically significant at $p<0.001$. With the lesser samples, the biases are a little worse.

The second result to draw from Table 5 is that the full IP weights do well. Even with the most limited sample (just web respondents, and only those who were regular internet users), only a few of the covariate means we consider exhibit a statistically significant bias, and the size of the biases are modest, at most 2 percentage points (for incidence of mortgage, smoking and long-standing illness).

Including those who are not regular internet users seems to be of little value. There are very limited changes in the biases. The bias in the incidence of long-standing illness is no longer statistically significant, but the bias in smoking incidence remains at 2 percentage points and is statistically significant at $p<0.01$. In contrast, including the telephone respondents does seem to have some value. With the telephone respondents and the IP weights, there are no biases in excess of 1 percentage point, and no biases that are statistically significant at $p<0.05$.

Table 6: COVID survey variables, inverse probability (IP) weighted means / prevalences and standard errors (brackets) for each cumulative dataset (Regular internet users, Web = regular internet users + non-regular internet users, Web and telephone).

Variable	Regular internet users	Web	Web and telephone
Advised to shield by NHS	0.07	0.08	0.09
	(0.00)	(0.01)	(0.01)
Reported suffering from asthma	0.15	0.15	0.14
	(0.01)	(0.01)	(0.01)
Reported suffering from arthritis	0.12	0.12	0.14
	(0.00)	(0.00)	(0.01)
Reported suffering from cancer	0.04	0.04	0.05
	(0.00)	(0.00)	(0.00)
In work	0.63	0.62	0.57
	(0.01)	(0.01)	(0.01)
Household net earnings (£/month)	1932	1905	1744
	(29.5)	(28.8)	(27.0)
On benefits	0.14	0.14	0.14
	(0.01)	(0.01)	(0.01)
Carer in own or other HH	0.47	0.46	0.44
	(0.01)	(0.01)	(0.01)

Finally, Table 6 shows how estimates of health and economic incidences during the pandemic (i.e., from the COVID Survey Wave 1 web and telephone surveys) vary with the sample employed. These are true survey targets, but we do not have a natural benchmark to compare these estimates to. Nevertheless, it is of interest to compare them to each other. These means are calculated using the full IP weights. We see modest but potentially important differences across the sample. For example, relative to the sample of only regular internet users responding by web, the broadest sample (including telephone respondents) estimates a significantly lower employment rate (57% versus 63%).

5. Discussion and Conclusions

In this paper we reported on the *Understanding Society* COVID-19 Study's experience with bias prevention and bias reduction. An attempt was made to prevent bias by inviting the full range of *Understanding Society* sample members - both those who regularly use the internet and those who do not - to participate and, furthermore, by conducting a telephone follow-up of non-respondents. An IP weighting strategy that takes advantage of the rich background information available from the Main Study was developed to reduce bias. In this example, weighting and the telephone follow-up helped to reduce bias, but inviting those who do not regularly use the internet to the web survey appears to have been of little benefit. Not only did inviting non-regular internet users offer little bias reduction, it also did not reduce the variability of the IP weights, suggesting that it also did not increase sample efficiency. In contrast, the additional respondents achieved through the telephone follow-up both reduced bias and weight variability. A further dimension of "representativeness" is providing sufficient samples of various subgroups of interest to support subgroup analysis (Benzeval *et al.*, 2020). While we have not analysed this explicitly, the very different demographics of telephone respondents (for example age; see Table 1) suggest that the telephone follow-up may be valuable in this way was well.

Against this, the telephone follow-up was costly. In the particular case of the COVID-19 Study, the average cost of a telephone response was 21.7 times the average cost of a web response, so that the total cost of the 718 telephone responses was almost as large as the total cost of the 17,761 web responses. There are fixed costs associated with both modes so that changes in scale or other parameters would affect the ratio of average costs. Similarly, had the two modes been conducted by the same fieldwork agency, some of the fixed costs might have been shared across modes. Nevertheless, the difference in costs is significant in reconsidering the design of such studies.

A further finding is that bias adjustments for web samples based on basic demographics alone do not seem to work well. The use of such simple adjustments with web samples is common. Our findings here echo the conclusions of (Couper *et al.*, 2007), despite the likelihood that patterns of internet use have evolved since that work. This suggests that population inferences from cross-sectional web surveys and non-probability samples should continue to be treated with caution.

Finally, we conclude that the COVID-19 Study provides a strong basis for population inferences. The combination of the telephone follow-up and IP weighting eliminated all biases in the covariates we considered. The COVID-19 Study was deployed rapidly, and the first data were released to researchers in less than three months from conception of the Study. Supplemental web surveys appear to be a highly valuable addition to existing panel studies in times of heightened data needs.

References

Adams-Prassl, A., Boneva, T., Golin, M., et al. (2020) Inequality in the impact of the coronavirus shock: Evidence from real time surveys. *Journal of Public Economics*, **189**, 104245. DOI: 10.1016/j.jpubeco.2020.104245.

Belot, M., Choi, S. and Tripodi, E. (2020) *Unequal Consequences of COVID-19 across Age and Income: Representative Evidence from Six Countries*. 14908, Discussion Paper. Centre for Economic Policy Research (CEPR).

Benzeval, M., Bollinger, C. R., Burton, J., et al. (2020) *The representativeness of Understanding Society*. 2020–08, Understanding Society Working Paper Series, March. Institute for Social and Economic Research. Available at: https://ideas.repec.org/p/ese/ukhlsp/2020-08.html.

Burton, J., Lynn, P. and Benzeval, M. (2020) How Understanding Society: The UK Household Longitudinal Study adapted to the COVID-19 pandemic. *Survey Research Methods*, , 235–239. Survey Research Methods. DOI: 10.18148/SRM/2020.V14I2.7746.

Couper, M. P., Kapteyn, A., Schonlau, M., et al. (2007) Noncoverage and nonresponse in an Internet survey. *Social Science Research*, **36**, 131–148. DOI: 10.1016/j.ssresearch.2005.10.002.

Crossley, T. F., Fisher, P. and Low, H. (2021) The Heterogeneous and Regressive Consequences of COVID-19: Evidence from High Quality Panel Data. *Journal of Public Economics*, **193**.

Denver, J. and Valliant, R. (2018) *Survey Weights: A Step-by-Step Guide to Calculation*. Stata Press.

Institute for Social and Economic Research (2020) Understanding Society COVID-19 User Guide Version 4.0. University of Essex.

Jäckle, A., Burton, J. and Couper, M. (2019) Event-trigged Data Collection. **2019**. Undestanding Society Working Paper.

Lynn, P. (2009) Sample Design for Understanding Society. *Understanding Society Working Paper*, **2009–01**, 46.

Lynn, P. and Kaminska, O. (2010) *Weighting Strategy for Understanding Society*. 2010–05, Understanding Society Working Paper Series, November. Institute for Social and Economic Research. Available at: https://ideas.repec.org/p/ese/ukhlsp/2010-05.html.

Moffit, R., Fitzgerald, J. and Gottschalk, P. (1999) Sample Attrition in Panel Data: The Role of Selection on Observables. *Annales d'Économie et de Statistique*, , 129–152. DOI: 10.2307/20076194.

Nicoletti, C. and Peracchi, F. (2005) Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *J Royal Statistical Soc A*, **168**, 763–781. DOI: 10.1111/j.1467-985X.2005.00369.x.

Schaurer, I. and Weiß, B. (2020) Investigating selection bias of online surveys on coronavirus-related behavioral outcomes. *Survey Research Methods*, **14**, 103–108. DOI: 10.18148/srm/2020.v14i2.7751.

Schonlau, M., Soest, A. van, Kapteyn, A., et al. (2009) Selection Bias in Web Surveys and the Use of Propensity Scores: *Sociological Methods & Research*. DOI: 10.1177/0049124108327128.

Schouten, B., Cobben, F., Lundquist, P., et al. (2016) Does more balanced survey response imply less non-response bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **179**, 727–748. DOI: 10.1111/rssa.12152.

Solon, G., Haider, S. J. and Wooldridge, J. M. (2015) What Are We Weighting For? *J. Human Resources*, **50**, 301–316. DOI: 10.3368/jhr.50.2.301.

University of Essex, I. for S. and E. R. (2020) Understanding Society: COVID-19 Study, 2020, 3rd Edition. DOI: 10.5255/UKDA-SN-8644-3.

University of Essex, I. for S. and E. R., NatCen Social Research and Kantar Public (2019) Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009, 12th Edition. DOI: 10.5255/UKDA-SN-6614-13.