Understanding Society

THE UK HOUSEHOLD LONGITUDINAL STUDY

Non-response bias prevention and adjustment in a national longitudinal Covid-19 survey

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

[1] Institute of Social and Economic Research, University of Essex
[2] European University Institute, Florence
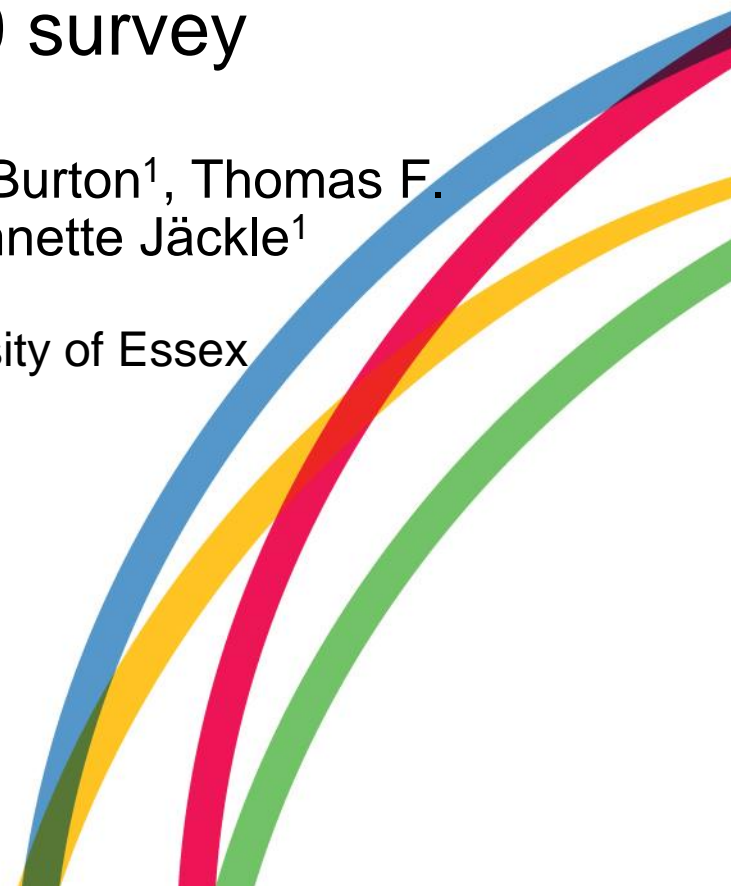[3] Ipsos MORI Social Research Institute

[*] Author contributions are equal, I'm presenting..

University of Essex

iiSER

UKRI

Economic and Social Research Council

The UKHLS main survey and Covid-19 Study

- UKHLS main survey a multi-domain longitudinal survey of the UK population.

- Started in 2009 with a probability sample of households.

- Yearly interviews, multi-mode, fixed and rotating items.

- More timely data needed during pandemic, so (bi-)monthly survey also funded.

- Covid-19 Study started in March 2020, (primarily) web based.

- Baseline items concerning Jan / Feb 2020 asked of all first time respondents, otherwise both fixed and rotating items.

- A major information source on impacts of Covid-19 in the UK: see https://www.understandingsociety.ac.uk/research/publications/subject/Covid%2019.

Minimising Covid-19 Study non-response errors

- Survey non-response common, tends to be higher with web interviewing.

- Increased by short data collection period in the Covid-19 Study?

- If differential across sample sub-groups can cause non-response error (survey estimate biases), reducing quality of inferences about the study population.

- Two ways of reducing biases, both which we sought to use in the Covid-19 Study:

     a) Bias prevention: methods used during data collection.

     b) Bias adjustment: methods used post data collection.

- In this talk, we describe these aspects of the survey design, and evaluate their performance.

Covid-19 Study bias prevention

- We sought to sample respondents with same attributes as the UK population by:

  - Using an eligible set of subjects in UKHLS wave 9 (2017/18) HHs.

  - Sending multiple reminders to non-respondents.

  - Telephone follow up of some non-regular internet users.

- Hence, three (partly overlapping) sample components:

  1) regular internet users issued the web survey.

  2) non-regular internet users issued the web survey.

  3) non-responding non-regular internet users issued to telephone.

- To evaluate performance, we quantify how respondent UKHLS w9 measured socio-demographic characteristics from the components in wave 1 compare to each other and to those of the eligible set.

Covid-19 Study sample sizes & wave 1 response rates

	All Eligible	Eligible set components			Respondents
	(i)	(ii)	(iii)	(iv)	(v)
		Regular Net Users	Non-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

For full results from the research in this talk, see Benzeval et al. (2021) *Understanding Society working paper 2021-03*.

Covid-19 Study sample & wave 1 respondent demographics

	All Eligible	Respondents			
	(i)	(ii)	(iii)	(iv)	(v)
		Regular net Users	Non-regular net Users	Issued to Tel. Survey	= (ii) + (iii) + 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: 40-49	0.16	0.19	0.10	0.05	0.17
Age: 70-79	0.12	0.12	0.16	0.30	0.13
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
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

Bias adjustment

- To reduce biases remaining after data collection, Inverse (Response) Propensity (IPW) weights scaling respondents to UK population.

- Adjusted UKHLS weights[*] i.e. response modelled conditional on w9 response.

- Probit model, UKHLS w9 predictors, *a priori* variable selection using Lasso.

- To evaluate performance, for Covid-19 Study w1 quantify weight variability with CVs & estimate variance inflation with Kish's DEFF.

- Plus, compare weighted UKHLS w9 variable estimates from (unbiased) UKHLS w9 and Covid-19 Study datasets.

- Use a novel statistical test[+] to compare estimates given partial sample dependence.

[*]For evidence that the UKHLS provides unbiased (weighted) estimates for the UK population, see Benzeval et al. (2020) *Understanding Society working paper 2020-08*.
[+]see Benzeval et al. (2021) *Understanding Society working paper 2021-03*.

Covid-19 Study wave 1 web survey Lasso selected predictors

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***

Covid-19 Study wave 1 weights: CVs & DEFFs

	(i)	(ii)	(iii)
	Regular internet users	All web	All web + telephone
		= (i) + non-regular internet users	= (ii) + telephone
DEFF	2.6	2.7	2.2
CV	140.1	156.7	133.6

Covid-19 Study wave 1 weights: UKHLS w9 variable biases

Variable	Wave 9 wt. est.	Regular net users IPW wt. diff.	Web IPW wt. diff.	Web and telephone IPW wt. diff
Subjective financial situation (SFS): comfortable or OK	0.71 (0.00)	0.00	0.00	0.00
SFS: just about getting by	0.21 (0.00)	0.00	0.00	-0.00
SFS: finding it quite/very difficult	0.07 (0.00)	-0.00	-0.01	-0.00
Tenure: Owned	0.34 (0.00)	0.01*	0.01	-0.00
Tenure: Mortgage	0.34 (0.00)	-0.02***	-0.01**	0.01
Tenure: Rented	0.13 (0.00)	-0.01	0.00	0.00
Tenure: Social Housing	0.19 (0.00)	0.01	0.00	-0.01
Low skill occupation	0.38 (0.01)	-0.01	-0.01	-0.01
Any savings income	0.36 (0.00)	-0.00	-0.01	0.00
Behind with some or all bills	0.06 (0.00)	0.00	0.00	0.00

Covid-19 Study wave 1 weights: UKHLS w9 variable biases

Variable	Wave 9 wt. est.	Regular net users IPW wt. diff.	Web IPW wt. diff.	Web and telephone IPW wt. diff
Not in IPW model:				
Income poverty	0.15	0.01	0.01	-0.01
	(0.00)			
Receives core benefit	0.05	-0.00	-0.00	-0.00
	(0.00)			
Behind with housing	0.09	0.00	0.00	-0.00
	(0.00)			
Smoker	0.15	0.02*	0.02**	0.01
	(0.00)			
Long-standing illness	0.37	0.02**	0.01	-0.01
	(0.00)			

Covid-19 Study wave 1 variable weighted estimates

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)

Summary

- Both bias prevention and adjustment used to reduce non-response errors (survey estimate biases) in the UKHLS Covid-19 Study.

- Evaluations of wave 1 dataset show that the strategy performs well.

- Telephone follow up (bias prevention) increases unweighted dataset quality compared to regular net users only.

- IPW weighting (bias adjustment) reduces remaining biases, most effectively in web plus telephone dataset.

- Interestingly, web sampled non-regular net users don't really improve quality.

  - small reduction in precision as well as biases.

- Could move straight to telephone, but *21.7 cost of web.

  - so ~700 sampled (~140 interviews), vs. ~2250 web interviews???

Contact

Jamie Moore:

moorej@essex.ac.uk

Comparing weighted estimates

- Construct the joint null that response to the two waves is independent of variable of interest given the auxiliary variables (pre-response observables):

$$E\left[R_{t-1,i}\middle|Y_{t-1,i},X_{t-2,i}\right] = E\left[R_{t-1,i}\middle|X_{t-2,i}\right] = {}^{1}\!/\!_{W_{t-1,i}}, \text{ and}$$

$$E\left[R_{t,i}\middle|Y_{t-1,i},X_{t-1,i}\right] = E\left[R_{t,i}\middle|X_{t-1,i}\right] = {}^{1}\!/\!_{W_{t,i}},$$

where:

$R_{t-1,i}$ = subject i response at main survey wave 9 (0 = No, 1 = Yes).

$R_{t,i}$ = subject i response at main survey wave 9 (0 = No, 1 = Yes).

$Y_{t-1,i}$ = subject i main survey wave 9 variable of interest value.

$X_{t-2,i}$ = subject i auxiliary variable (pre-response observable) values at main survey wave 8.

$X_{t-1,i}$ = subject i auxiliary variable values at main survey wave 9.

$W_{t-1,i}$ = subject i main survey wave 9 non-response weight.

$W_{t,i}$ = subject i Covid-19 Study wave 1 non-response weight.

Comparing weighted estimates

- A test of the adequacy of the Covid-19 Study wave 1 weights is then provided by the moment condition:

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

where:

$s_{t-1,i}$ = is the subject i main survey w9 weight share = $W_{t-1,i}/\sum_{i=1}W_{t-1,i}$

$s_{t,i}$ = is the subject i Covid-19 Study wave 1 weight share = $W_{t,i}/\sum_{i=1}W_{t,i}$

- Condition captures fact that under the joint null either set of responses and associated weights gives a consistent estimate of $E(Y_{t-1})$.
- Summed over subjects and compared to T distribution with 1 d.f.