

# Handling non-response in COVID-19 surveys across five national longitudinal studies

Richard Silverwood & George Ploubidis  
UCL Centre for Longitudinal Studies

CENTRE FOR  
LONGITUDINAL  
STUDIES



Missing Data in Longitudinal and Linked Surveys  
11 December 2020



Economic  
and Social  
Research Council

# Outline

1. 1958 National Child Development Study
2. COVID-19 surveys
3. Non-response weight derivation
4. Effectiveness of non-response weights
5. Alternative approaches for handling non-response

# 1958 National Child Development Study

# Five national longitudinal studies

1946 National Study of Health  
and Development



**1958 National Child Development Study**



1970 British Cohort Study



Next Steps (formerly LSYPE)








Millennium Cohort Study






# 1958 National Child Development Study (NCDS)

- Longitudinal birth cohort study of all babies born in a single week in Great Britain in 1958.
- Initial N = 17,415, later augmented by immigrants born in target week.
- Multidisciplinary content spanning physical and educational development, economic circumstances, employment, family life, health behaviour, wellbeing, social participation and attitudes.

# NCDS: Respondents, instruments and response

	1958 Birth	1965 7	1969 11	1974 16	1981 23	1991 33	2000 42	2003 44	2004 46	2008 50	2013 55
 main respondent	mother	parents	parents	subject/ parents	subject	subject	subject	subject	subject	subject	subject
 secondary respondent	medical	medical/ school	medical/ school	medical/ school		partner mother children					
 survey instruments		cognitive assess- ments	cognitive assess- ments	cognitive assess- ments				Bio- medical data collection		cognitive assess- ments	
 linked data				area of residence (census)	area of residence (census)						
 response	17,415	15,425	15,337	14,654	12,537	11,469	11,419	9,377	9,534	9,790	9,137

# NCDS: Topics covered by life stage

 Birth	 School years	 Adult
<p>Family</p> <p>Parental employment</p> <p>Obstetric history</p> <p>Smoking in pregnancy</p> <p>Pregnancy (problems, antenatal care)</p> <p>Labour (length, pain relief, problems)</p> <p>Birth (problems, weight, gest age)</p>	<p>Family</p> <p>Parental employment</p> <p>Financial circumstances</p> <p>School</p> <p>Housing</p> <p>Views and expectations</p> <p>Attainment</p> <p>Health</p> <p>Behaviour</p> <p>Cognition</p>	<p>Family (partners, children)</p> <p>Employment</p> <p>Income</p> <p>Housing</p> <p>Courses and qualifications</p> <p>Basic skills</p> <p>Views and expectations</p> <p>Health</p> <p>Health-related behaviour</p> <p>Cognition</p>





# Cross-cohort COVID-19 surveys

Study	Age in 2020
MCS	19
Next Steps	30
BCS70	50
<b>NCDS</b>	<b>62</b>
NSHD	74

# Cross-cohort COVID-19 surveys: Wave 1

- Online survey in May 2020 (height of lockdown restrictions).
- Focused mainly on how lives had changed from just before the outbreak of the pandemic.
- Topics included:
  - Physical health (including COVID-19)
  - Time use
  - Family and household
  - Financial situation and benefits
  - Employment and education
  - Health behaviours
  - Mental health and social connectedness
- Over 18,000 participants. Data available via UK Data Service.

# Cross-cohort COVID-19 surveys: Wave 2

- Online survey in September/October 2020.
- Aimed to capture how lives have changed from Wave 1.
- Topic areas mirror closely those for Wave 1, with additional questions about:
  - Health care
  - Financial transfers
  - Life events
  - Children's schooling in summer and autumn term
- Almost 26,000 participants. Data available via UK Data Service imminently.
- (Wave 3 planned for early 2021.)



# Non-response

- Non-response is common in longitudinal surveys.
- Missing values mean less efficient estimates because of reduced size of analysis sample.
- Also introduces potential for bias since respondents are often systematically different from non-respondents.
- Well known methods for dealing with missing data include multiple imputation (MI), inverse probability weighting (IPW), and full information maximum likelihood (FIML).

## Non-response weights

- To correct for non-response in the COVID-19 surveys, non-response weights are provided, so that IPW analyses can be undertaken.
- Non-response weights capitalise on the rich data cohort members have provided over many years.

## Target population

- Target population of each cohort defined as individuals born in the specified birth period who are alive and still residing in the UK.
- Non-response weights designed to make weighted results from COVID-19 survey respondents representative of the target population.
- COVID-19 surveys also issued to a relatively small number of cohort members who had already emigrated from the UK – we do not derive non-response weights for such individuals.

# Response

Cohort	Wave 2			
	Issued sample	Response within issued sample	Target population	Response within target population
NSHD	2551	1569 (61.5%)	3758	1488 (39.6%)
<b>NCDS</b>	<b>11,655</b>	<b>6282 (53.9%)</b>	<b>15,291</b>	<b>6228 (40.7%)</b>
BCS70	12,133	5320 (43.9%)	17,486	5236 (29.9%)
Next Steps	11,529	3664 (31.8%)	15,770	3609 (22.9%)
MCS cohort members	13,547	3274 (24.2%)	19,243	3233 (16.8%)
Total	51,415	20,109 (39.1%)	71,548	19,794 (27.7%)

Overall response rate within issued sample (39.1%) comparable to similar COVID-19 web surveys (e.g. Understanding Society COVID-19 Web Survey Wave 4: 38.0%).



# Derivation of non-response weights

## Overview

1. Within sample corresponding to target population, model COVID-19 survey response conditional on a common set of covariates using logistic regression.
2. For COVID-19 survey respondents, predict probability of response from model.
3. Calculate non-response weight as inverse of probability of response.
4. Examine distribution of weights across cohorts to decide whether truncation may be desirable; apply truncation if so.
5. Calibrate weights so they sum to number of respondents in each cohort.

# Derivation of non-response weights

## Stage 1: Response model

- Selection of covariates in response model informed by literature and results of the CLS Missing Data Strategy, plus assumed associations with the probability of response and/or with key COVID-19 survey variables.
- Aimed to use broadly same set of variables in each cohort to ensure consistency.
- Not possible to include identical sets of variables due to data being collected at different ages and using different questions.

# Derivation of non-response weights

## Stage 1: Response model

Sex  
Ethnicity  
Parental social class  
Number of rooms at home/persons per room  
Cognitive ability  
Early life mental health  
Voting  
Membership in organisations

Internet access prior to web survey  
Consent for biomarkers  
Consent for linkages  
Educational qualifications  
Economic activity  
Partnership status  
Psychological distress  
BMI

Self-rated health  
Smoking status  
Maternal mental health  
Social capital/social support  
Income  
Number of non-responses across all previous sweeps  
Response at COVID-19 Wave 1 survey\*

# Derivation of non-response weights

## Stage 1: Response model

- Missing covariate values handled using multiple imputation (MI), conducted in each cohort separately.
- Imputation model included above variables, COVID-19 Wave 1 (and Wave 2\*) survey response and, for relevant cohorts (NSHD, Next Steps and MCS), the design weight.
- Five imputed datasets were created using chained equations.
- Models for COVID-19 survey response were fitted in each imputed dataset and combined using standard rules.

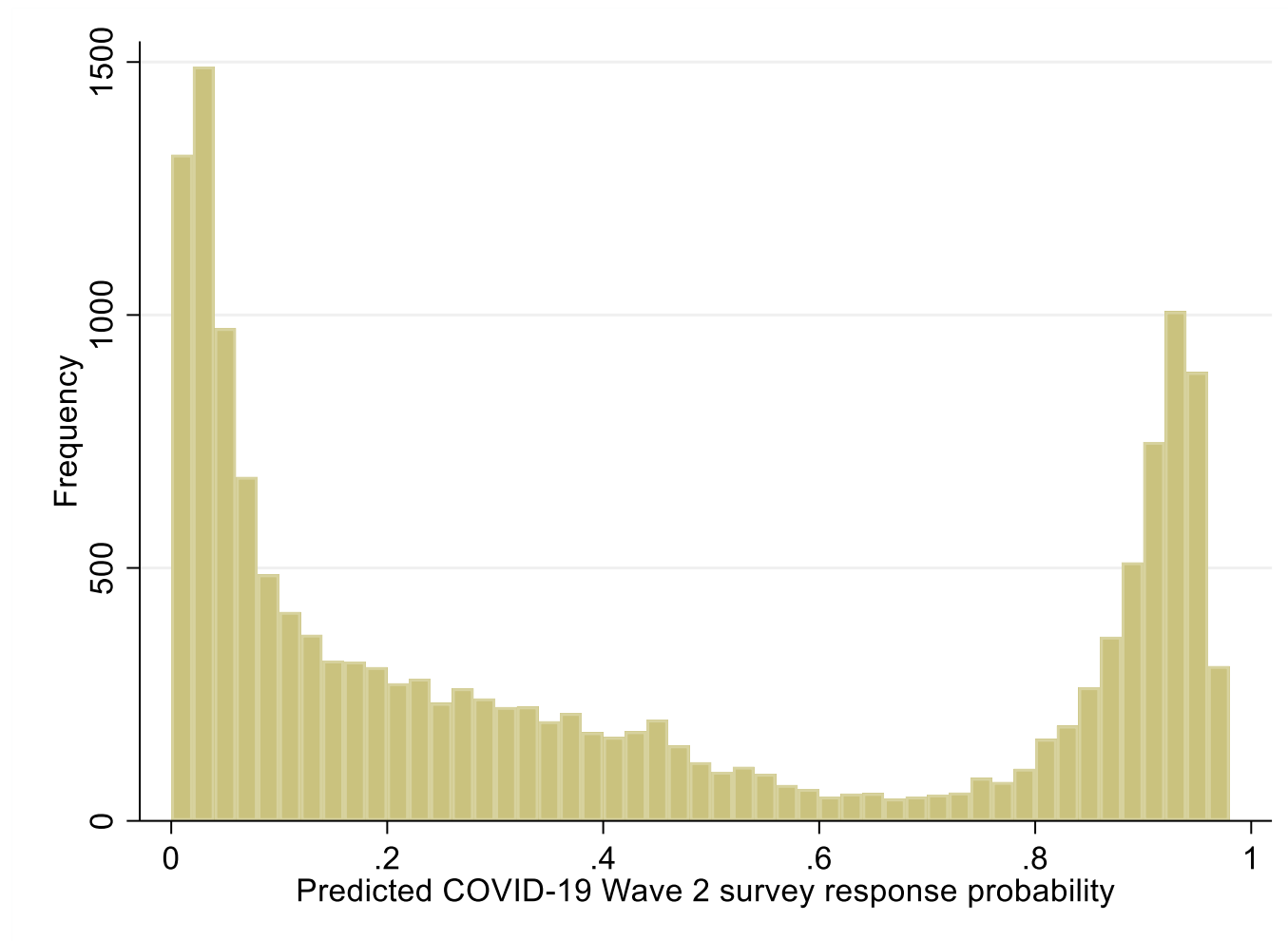
# Derivation of non-response weights

## Stage 2: Predict probability of response

1. Within sample corresponding to target population, model COVID-19 survey response conditional on a common set of covariates using logistic regression.
2. For COVID-19 survey respondents, predict probability of response from model.
3. Calculate non-response weight as inverse of probability of response.
4. Examine distribution of weights across cohorts to decide whether truncation may be desirable; apply truncation if so.
5. Calibrate weights so they sum to number of respondents in each cohort.

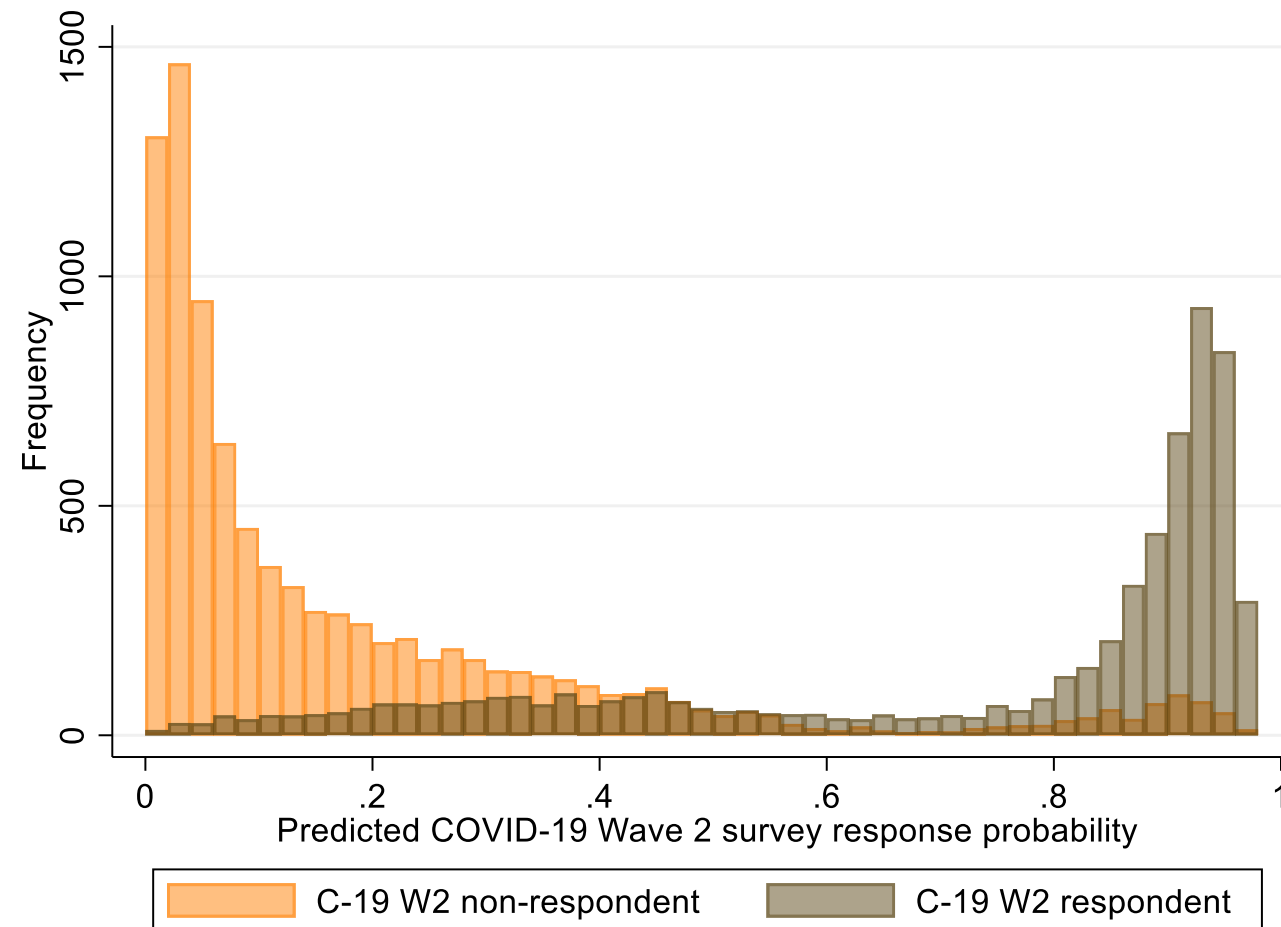
# Derivation of non-response weights

## Stage 2: Predict probability of response



# Derivation of non-response weights

## Stage 2: Predict probability of response



# Derivation of non-response weights

## Stage 2: Predict probability of response

Predicted COVID-19 Wave 2 survey response probability	n	COVID-19 Wave 2 survey respondents (%)
[0.0, 0.1)	4,949	143 (2.9)
[0.1, 0.2)	1,717	243 (14.2)
[0.2, 0.3)	1,290	354 (27.4)
[0.3, 0.4)	1,036	394 (38.0)
[0.4, 0.5)	808	391 (48.4)
[0.5, 0.6)	431	248 (57.5)
[0.6, 0.7)	249	193 (77.5)
[0.7, 0.8)	373	285 (76.4)
[0.8, 0.9)	1,489	1,254 (84.2)
[0.9, 1.0]	2,949	2,723 (92.3)
Total	15,291	6,228



# Derivation of non-response weights

## Stages 2-5

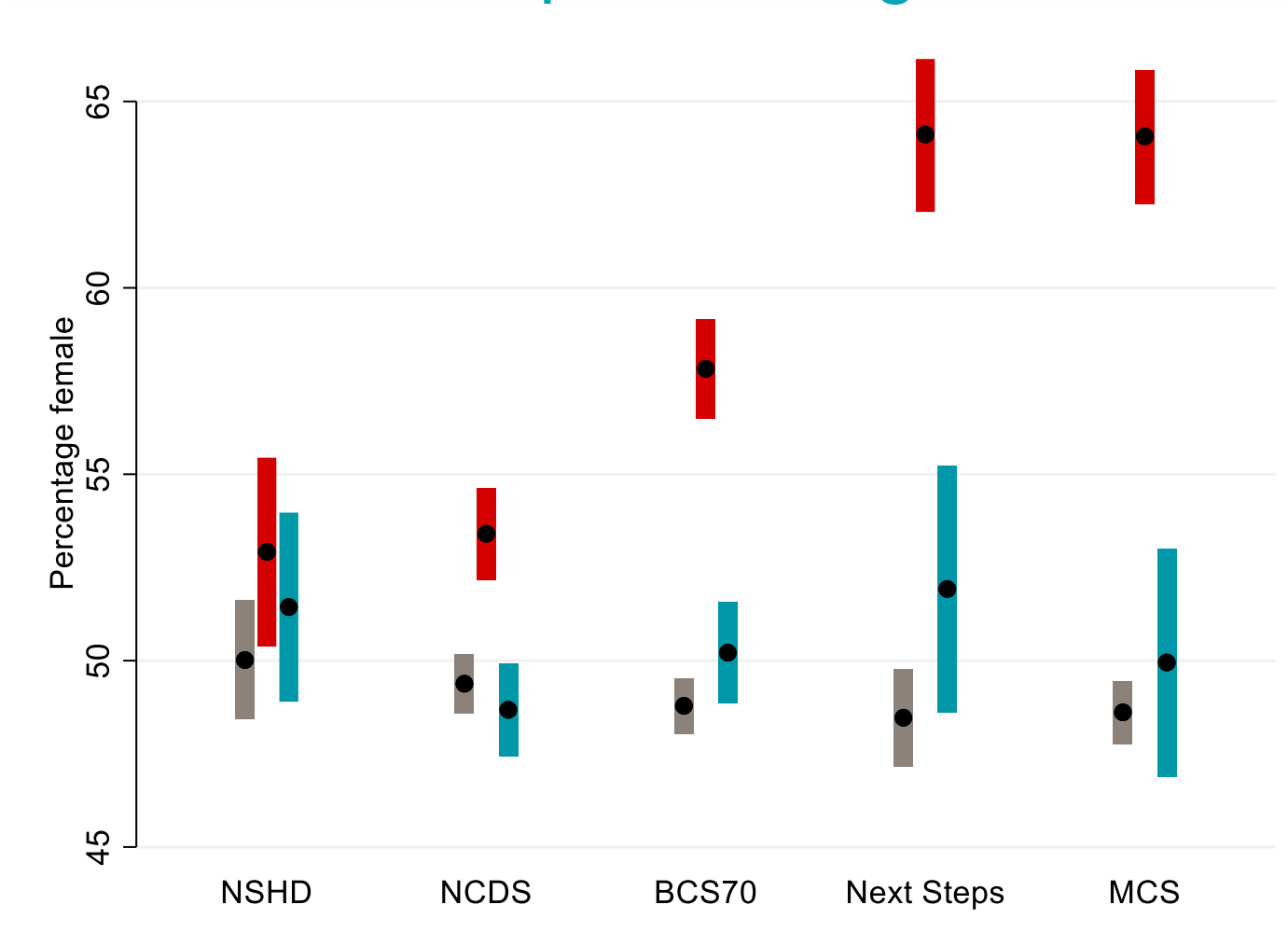
1. Within sample corresponding to target population, model COVID-19 survey response conditional on a common set of covariates using logistic regression.
2. For COVID-19 survey respondents, predict probability of response from model.
3. Calculate non-response weight as inverse of probability of response.
4. Examine distribution of weights across cohorts to decide whether truncation may be desirable; apply truncation if so.
5. Calibrate weights so they sum to number of respondents in each cohort.



# Effectiveness of non-response weights

- To examine effectiveness of non-response weights in restoring sample representativeness we conducted several analyses.
- We considered the distribution of sex in each cohort, which is observed at baseline in virtually all cohort members.
- We compared the distribution of sex:
  - In all cohort members;
  - In COVID-19 survey respondents only;
  - In COVID-19 survey respondents after application of non-response weights.

# Effectiveness of non-response weights



Grey: all cohort members; red: COVID-19 Wave 2 survey respondents only; blue: COVID-19 Wave 2 survey respondents after application of non-response weights.



# Alternative approaches for handling non-response

- NCDS only.
- Are people already in worse health more likely to contract COVID-19?
- Self-reported general health at age 50 obtained from responses to the question:

In general, would you say your health is...?

- Excellent
- Very good
- Good
- Fair
- Poor

# Alternative approaches for handling non-response

- COVID-19 status at the Wave 2 survey obtained from responses to the question:  
Do you think that you have or have had Coronavirus?
  - Yes, confirmed by a positive test
  - Yes, based on strong personal suspicion or medical advice
  - Unsure
  - No
- Logistic regression, adjusted for sex.

## COVID-19 status and prior health

- Target population: 15,291.
- COVID-19 Wave 2 survey respondents: 6,228.
- Non-missing for COVID-19 status: 6,185.
- Non-missing for self-rated health at age 50: 9,477.
- Non-missing for sex: 15,291.
- Non-missing for sex, self-rated health and COVID-19 status: 5,622.



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

# COVID-19 status and prior health

	Weight derivation	Derived	Imputation	Analysis		
	S					
IPW-GEN-MI						

- Inverse probability weighted (**IPW**) analysis using generic (**GEN**) multiple imputation (**MI**)-derived weights.
- Generic = modelling unit response.
- $M = 5$ .

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

COVID-19 Wave 2 survey respondents.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

Complete data on all analysis variables.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>						
<b>IPW-GEN-MI</b>	10,251	0,220		0,022	1.00	0.00, 1.01

Analysis using observed (**OBS**) data.



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

Complete data on all analysis variables.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS					5.000	1.40 1.00 1.00
IPW-GEN-MI						
IPW-AS-MI						

- Inverse probability weighted (**IPW**) analysis using analysis-specific (**AS**) multiple imputation (**MI**)-derived weights.
- Analysis-specific = modelling CC probability.
- Imputation model includes CC indicator rather than COVID-19 W2 survey response indicator.
- $M = 5$ .



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75

Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75

Complete data on all analysis variables.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75

Complete data on all analysis variables.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77

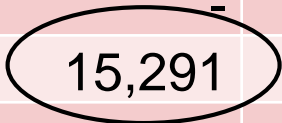
# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS	-	-	-	5,622	1.42	1.08, 1.86
IPW-GEN-MI						
IPW-AS-MI						
MI-GEN-ALL						

- Multiple imputation (**MI**) analysis of all cohort members (**ALL**) using generic (**GEN**) imputation model.
- Same variables as weight derivation imputation model, except include COVID-19 status rather than COVID-19 Wave 2 survey response indicator.
- M = 50.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77



Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77

Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81

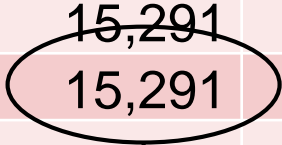
# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS	-	-	-	5,622	1.42	1.08, 1.86
IPW-GEN-MI						
IPW-AS-MI						
MI-GEN-ALL						
MI-GEN-ID						

- Multiple imputation (**MI**) analysis using generic (**GEN**) imputation model, with outcome variable imputed then deleted (**ID**).
- Same imputation model as for MI-GEN-ALL.
- $M = 50$ .

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81



Whole sample.



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81

Non-missing for COVID-19 status.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81

# COVID-19 status and prior health

	Weight derivation	Derived	Imputation	Analysis		
OE						6
IPV						1
IPV						5
MI						7
MI						1

Which variables should be included in the imputation model?

- All the variables in the substantive (analysis) model (including the outcome).
- Auxiliary variables (not in the substantive model):
  - Variables predictive of both the chance of missing values and the underlying values themselves.
  - Variables predictive of only the underlying missing values.
  - (Not variables only predictive of the chance of missing values).

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS	-	-	-	5,622	1.42	1.08, 1.86
IPW-GEN-MI	15,291	6,228	-	5,622	1.33	0.93, 1.91
IPW-AS-MI	15,291	5,622	-	5,622	1.44	0.74, 1.75
MI-GEN-ALL						
MI-GEN-ID						
MI-AS-ALL						

- Multiple imputation (**MI**) analysis of all cohort members (**ALL**) using analysis-specific (**AS**) imputation model.
- Imputation model includes self-rated health at ages 33, 42 and 46.
- M = 50.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73

Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73

Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS	-	-	-	5,622	1.42	1.08, 1.86
IPW-GEN-MI						
IPW-AS-MI						
MI-GEN-ALL						
MI-GEN-ID						
MI-AS-ALL						
MI-AS-ID						

- Multiple imputation (**MI**) analysis using analysis-specific (**AS**) imputation model, with outcome variable imputed then deleted (**ID**).
- Imputation model same as for MI-AS-ALL.
- M = 50.



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81

Whole sample.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81

Non-missing for COVID-19 status.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81
<b>IPW+MI-GEN-ID</b>	15,291	6,228	15,291	6,185	1.35	0.61, 3.01

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS					5.622	4.42 - 7.09
IPW-GEN-MI						
IPW-AS-MI						
MI-GEN-ALL						
MI-GEN-ID						
MI-AS-ALL						
MI-AS-ID						
IPW+MI-GEN-ID						

- Multiple imputation (**MI**) analysis using generic (**GEN**) imputation model, with outcome variable imputed then deleted (**ID**), inverse probability weighted (**IPW**) using generic MI-derived weights
- Same imputation model as MI-GEN-ID.
- M = 50.

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81
<b>IPW+MI-GEN-ID</b>	15,291	6,228	15,291	6,185	1.35	0.61, 3.01



# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81
<b>IPW+MI-GEN-ID</b>	15,291	6,228	15,291	6,185	1.35	0.61, 3.01

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81
<b>IPW+MI-GEN-ID</b>	15,291	6,228	15,291	6,185	1.35	0.61, 3.01
<b>IPW+MI-AS-ID</b>	15,291	6,228	15,291	6,185	1.38	0.64, 3.00

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
OBS					5.622	4.42 - 7.08
IPW-GEN-MI						
IPW-AS-MI						
MI-GEN-ALL						
MI-GEN-ID						
MI-AS-ALL						
MI-AS-ID						
IPW+MI-GEN-ID						
IPW+MI-AS-ID						

- Multiple imputation (**MI**) analysis using analysis-specific (**AS**) imputation model, with outcome variable imputed then deleted (**ID**), inverse probability weighted (**IPW**) using generic MI-derived weights
- Same imputation model as MI-AS-ID.
- M = 50.

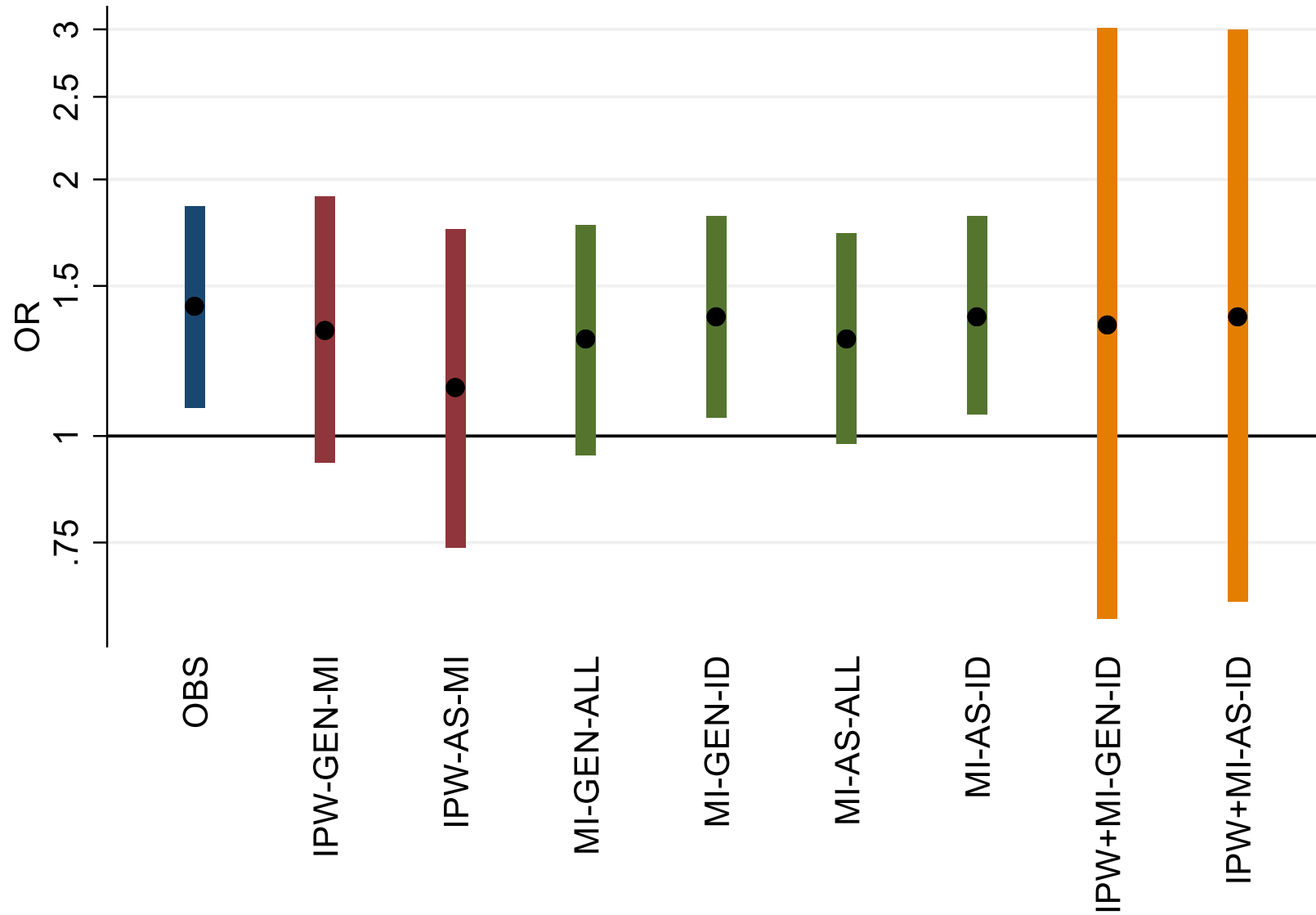
# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81
<b>IPW+MI-GEN-ID</b>	15,291	6,228	15,291	6,185	1.35	0.61, 3.01
<b>IPW+MI-AS-ID</b>	15,291	6,228	15,291	6,185	1.38	0.64, 3.00

# COVID-19 status and prior health

	Weight derivation sample	Derived weights	Imputation sample	Analysis sample	OR	95% CI
<b>OBS</b>	-	-	-	5,622	1.42	1.08, 1.86
<b>IPW-GEN-MI</b>	15,291	6,228	-	5,622	1.33	0.93, 1.91
<b>IPW-AS-MI</b>	15,291	5,622	-	5,622	1.14	0.74, 1.75
<b>MI-GEN-ALL</b>	-	-	15,291	15,291	1.30	0.95, 1.77
<b>MI-GEN-ID</b>	-	-	15,291	6,185	1.38	1.05, 1.81
<b>MI-AS-ALL</b>	-	-	15,291	15,291	1.30	0.98, 1.73
<b>MI-AS-ID</b>	-	-	15,291	6,185	1.38	1.06, 1.81
<b>IPW+MI-GEN-ID</b>	15,291	6,228	15,291	6,185	1.35	0.61, 3.01
<b>IPW+MI-AS-ID</b>	15,291	6,228	15,291	6,185	1.38	0.64, 3.00

# COVID-19 status and prior health



# Simulation study

- Focus is on how to handle missing data in both covariates and outcome with emphasis on real-world cohort application throughout.
- Simulate data based on real analysis of NCDS data for a realistic sample size: covariates, outcome and auxiliary variables all subject to missingness.
- Vary:
  1. Extent of missingness of covariates, outcome and auxiliary variables (varied separately).
  2. Magnitudes of associations between variables.
  3. Magnitudes of associations between variables and missingness.
- Compare above methods: bias, coverage properties.

# Simulation study

Questions of particular interest:

- In what circumstances do generic weights (derived using unit non-response) become unreliable?
- How strongly correlated with the outcome does an auxiliary variable have to be before you can safely analyse imputed outcome values?
- Can MI and IPW be combined more effectively?



## Further issues...

- Weight truncation: Deposited NCDS weights truncated (prior to calibration) so not identical to those used here.
- Appropriately accounting for complex sampling (clustering, stratification) in MI (→ multilevel MI?).
- MI uncertainty in MI-derived weights is not appropriately propagated into the IPW estimation.
- Methods considered only valid under the assumption of MAR. Missingness of COVID-19 status presumably may be affected by underlying COVID-19 status → MNAR. Could use MNAR-specific approach or utilise MNAR sensitivity analyses to explore impact of deviations from MAR assumption.

# Resources

- Missing data resources: <https://cls.ucl.ac.uk/data-access-training/handling-missing-data/>
- COVID-19 data: <https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8658>

Centre for Longitudinal Studies (CLS) website page titled "Handling missing data". The page includes a navigation menu with "COVID-19", "Our studies", "Our research", "Publications and resources", and "Data access and training". The main content area discusses the challenges of missing data in longitudinal studies and provides advice on handling it, including methods like multiple imputation and inverse probability weighting. A reference to Mostafa et al. [1] is provided at the bottom.

UK Data Service website page for study ID 8658. The page displays the study title: "COVID-19 Survey in Five National Longitudinal Cohort Studies: Millennium Cohort Study, Next Steps, 1970 British Cohort Study and 1958 National Child Development Study, 2020". It includes a "Copy study DOI" button and a "Details" section with the following information:

Title:	COVID-19 Survey in Five National Longitudinal Cohort Studies: Millennium Cohort Study, Next Steps, 1970 British Cohort Study and 1958 National Child Development Study, 2020
Study number (SN):	8658
Access:	These data are <b>safeguarded</b>
Persistent identifier	10.5255/UKDA-SN-8658-1

# Resources

- Mostafa T, Narayanan M, Pongiglione B, Dodgeon B, Goodman A, Silverwood RJ, Ploubidis GB. *Improving the plausibility of the missing at random assumption in the 1958 British birth cohort: A pragmatic data driven approach*. CLS Working Paper 2020/6. London: UCL Centre for Longitudinal Studies; 2020.
- Silverwood RJ, Calderwood L, Sakshaug JW, Ploubidis GB. *A data driven approach to understanding and handling non-response in the Next Steps cohort*. CLS Working Paper 2020/5. London: UCL Centre for Longitudinal Studies; 2020.
- Silverwood R, Narayanan M, Dodgeon B., Ploubidis G. *Handling missing data in the National Child Development Study: User guide*. London: UCL Centre for Longitudinal Studies; 2020.
- Brown M, Goodman A, Peters A, Ploubidis GB, Sanchez A, Silverwood R, Smith K. *COVID-19 Survey in Five National Longitudinal Studies: Waves 1 and 2 User Guide (Version 2)*. London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing; 2020.

# References

- Carpenter JR, Kenward MG. *Multiple Imputation and its Application*. Chichester, UK: John Wiley & Sons; 2013.
- Von Hippel PT. *Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data*. *Sociological Methodology*. 2007; 37(1): 83-117.
- Seaman SR, White IR, Copas AJ, Li L. *Combining multiple imputation and inverse-probability weighting*. *Biometrics*. 2012; 68(1): 129-37.

Thank you.