Institute of Education

CLS Missing Data Strategy

Richard Silverwood & Michalis Katsoulis

27 April 2023

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

Outline

CENTRE FOR LONGITUDINAL STUDIES

1. CLS Missing Data Strategy
2. 1958 National Child Development Study (NCDS)
3. 1970 British Cohort Study (BCS70)
4. Other CLS studies
5. COVID-19 Surveys
6. Leveraging data linkages
7. Resources

CENTRE FOR LONGITUDINAL STUDIES

CLS Missing Data Strategy

CLS Missing Data Strategy

CENTRE FOR LONGITUDINAL STUDIES

- All users of longitudinal data need to consider the issue of missing data since some non-response is inevitable.
- Strategies for how to deal with missing data depend on the nature of non-response.
- Well known (principled) methods for handling missing data include multiple imputation, inverse probability weighting and full information maximum likelihood.
- These rely on the assumption that the data are missing at random (MAR), implying that systematic differences between the missing values and the observed values can be explained by observed data.

CLS Missing Data Strategy

CENTRE FOR LONGITUDINAL STUDIES

- Most studies employing such MAR methods rely on a largely arbitrary selection of variables used as predictors of missingness.
- We aim to maximise the plausibility of the MAR assumption by optimising the set of such variables used in analyses.
- We use systematic (data driven) approaches to identify variables that are associated with non-response at each sweep in each study.
- This allows us to capitalise on the rich data cohort members have provided over the years/decades in order to deal with missing data and reduce bias.

Why the focus on (wave) non-response?

CENTRE FOR LONGITUDINAL STUDIES

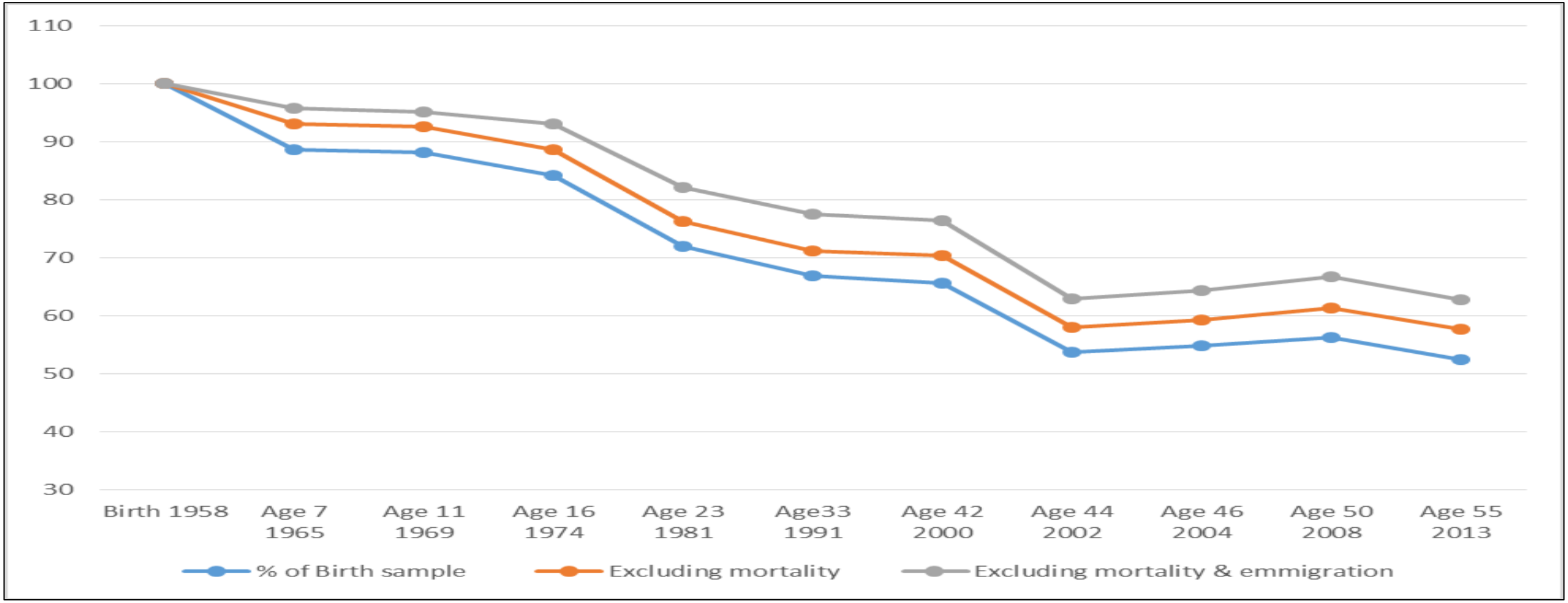
- Wave non-response is the main driver of missing data in analyses of CLS studies. Item non-response less of an issue.
- Much of the wave non-response is due to attrition.
- For longitudinal analyses, wave non-response at the most recent sweep is therefore usually the biggest contributor to missingness.
- Can identify predictors of wave non-response at cohort (rather than analysis) level – pragmatic approach.
- In analyses in which item non-response is more prevalent, this may need additional consideration.

CENTRE FOR LONGITUDINAL STUDIES

1958 National Child Development Study (NCDS)

Response in NCDS

CENTRE FOR LONGITUDINAL STUDIES



Identifying predictors of non-response in NCDS

CENTRE FOR LONGITUDINAL STUDIES

- Aim to maximise the plausibility of the MAR assumption by exploiting the richness of NCDS data.
- Using a data driven approach we identify the variables that are associated with non-response at each sweep.
- These can then be used as auxiliary variables.
- Substantive interest in understanding the drivers of non-response within and between cohorts.

Identifying predictors of non-response in NCDS

CENTRE FOR LONGITUDINAL STUDIES

- ~17,500 variables in NCDS sweeps 0-8.
- Exclude:
  - Routed variables.
  - Binary variables with prevalence <1%.
  - Variables with item non-response > 50%.
- Use summary scores for scales.
- Use summary measures; exclude constituent variables.
- 587 variables meeting inclusion criteria

  → multi-stage, data driven approach.

Identifying predictors of non-response in NCDS

CENTRE FOR LONGITUDINAL STUDIES

- For non-response at sweep *t*:
  - Stage 1: Univariable regressions for predictors at sweep 0, …, sweep *t* – 1. Retain predictors with $p < 0.05$.
  - Stage 2: Multivariable regressions for predictors at sweep 0, …, sweep *t* – 1. Retain predictors with $p < 0.05$.
  - Stage 3: MI. Multivariable regressions for predictors at sweep 0, …, sweep *t* – 1, adjusted for predictors at previous waves. Retain predictors with $p < 0.001$.
- Full details in Mostafa et al (2021) and NCDS Missing Data User Guide.

Predictors of non-response

CENTRE FOR LONGITUDINAL STUDIES

	NR sweep 1 (age 7)	NR sweep 2 (age 11)	NR sweep 3 (age 16)	NR sweep 4 (age 23)	NR sweep 5 (age 33)	NR sweep 6 (age 42)	NR BM sweep (age 44)	NR sweep 7 (age 46)	NR sweep 8 (age 50)	NR sweep 9 (age 55)
Sweep 0 (birth)	3	1	1	4	3	3	5	3	3	6
Sweep 1 (age 7)		5	3	3	5	1	5	4	3	4
Sweep 2 (age 11)			1	4	3	3	1	3	2	2
Sweep 3 (age 16)				4	4	3	4	4	4	5
Sweep 4 (age 23)					5	2	1	2	3	2
Sweep 5 (age 33)						5	4	2	3	5
Sweep 6 (age 42)							5	3	5	2
BM sweep (age 44)								3	3	1
Sweep 7 (age 46)									1	1
Sweep 8 (age 50)										3
Total	**3**	**6**	**5**	**15**	**20**	**17**	**25**	**24**	**27**	**31**

Consistent predictors of participation

CENTRE FOR LONGITUDINAL STUDIES

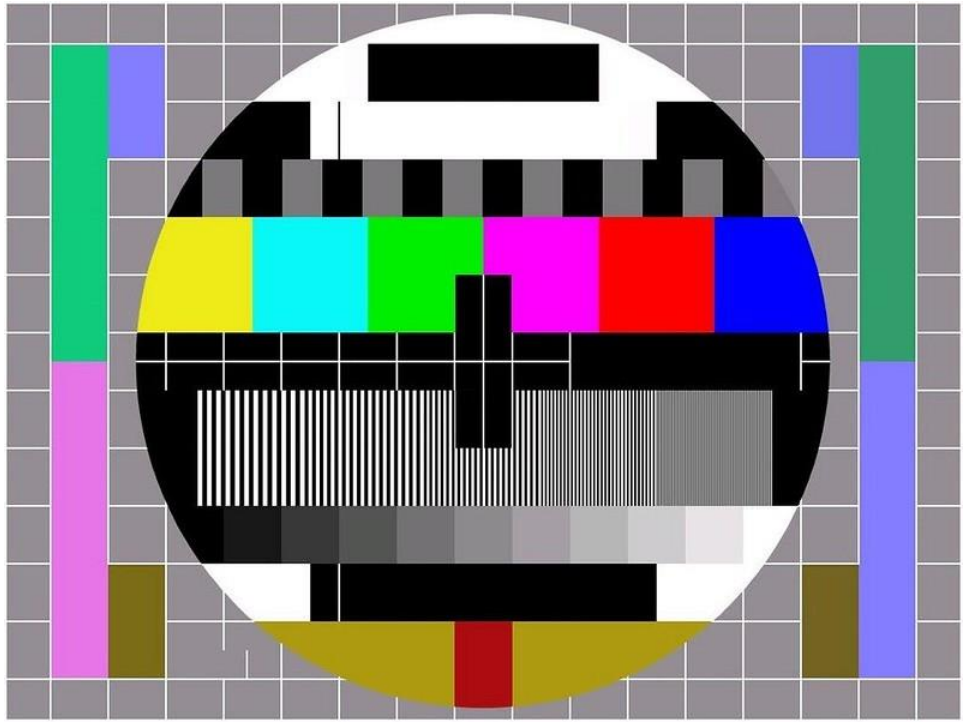
- Being female (adulthood only).
- Lower childhood social class in childhood; higher childhood social class in adulthood.
- Higher early life cognitive ability; fewer adolescent conduct problems.
- Social participation; voting; union membership.
- Being married; home ownership.
- Participation in previous sweeps.

Simple test analyses

CENTRE FOR LONGITUDINAL STUDIES

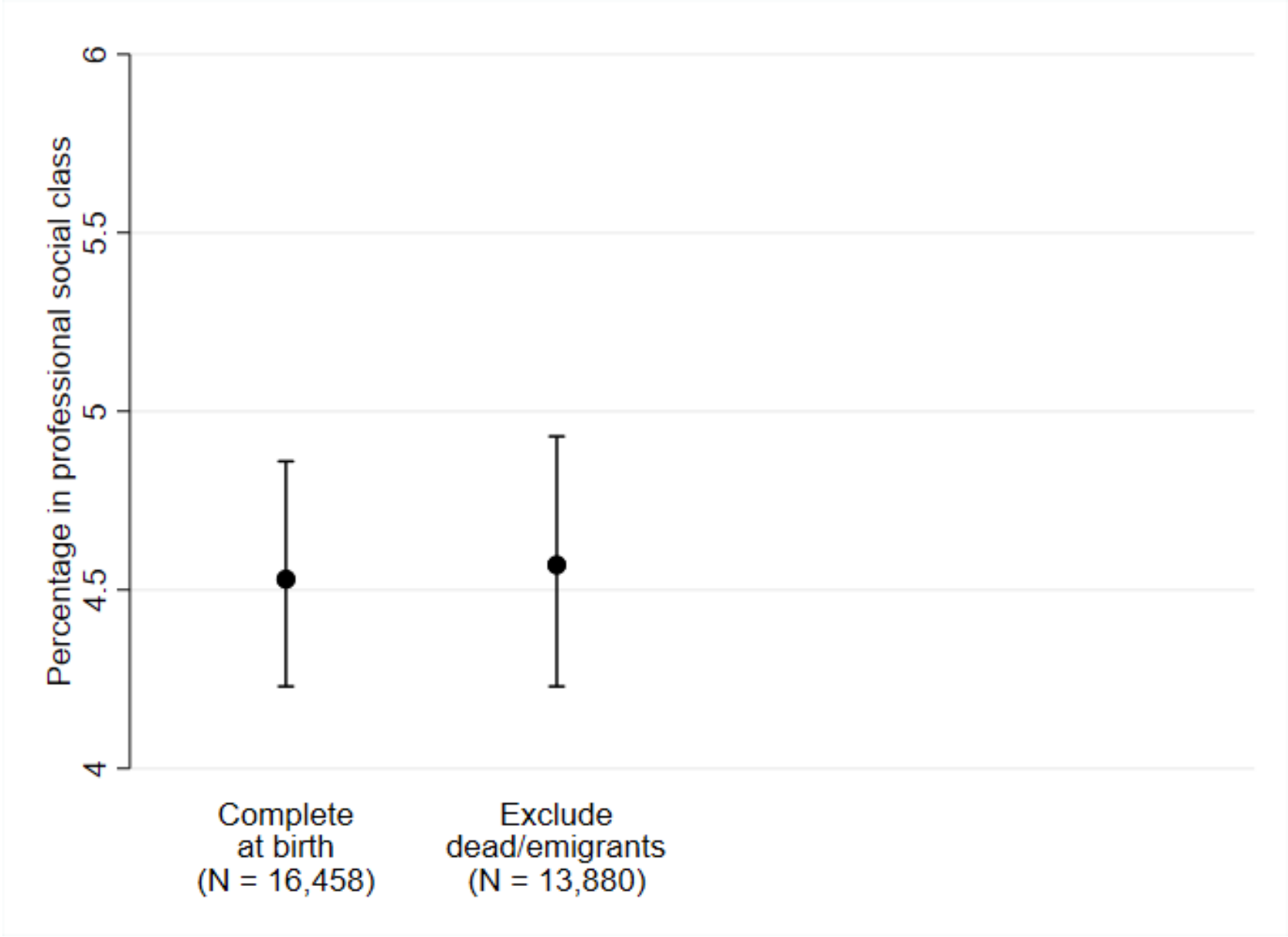
We test the performance of the missing data strategy using two approaches:

1. “Travelling back in time” to see whether distributions of variables from earlier sweeps can be replicated using only data from respondents at a later sweep.
2. Comparison to external population benchmarks.



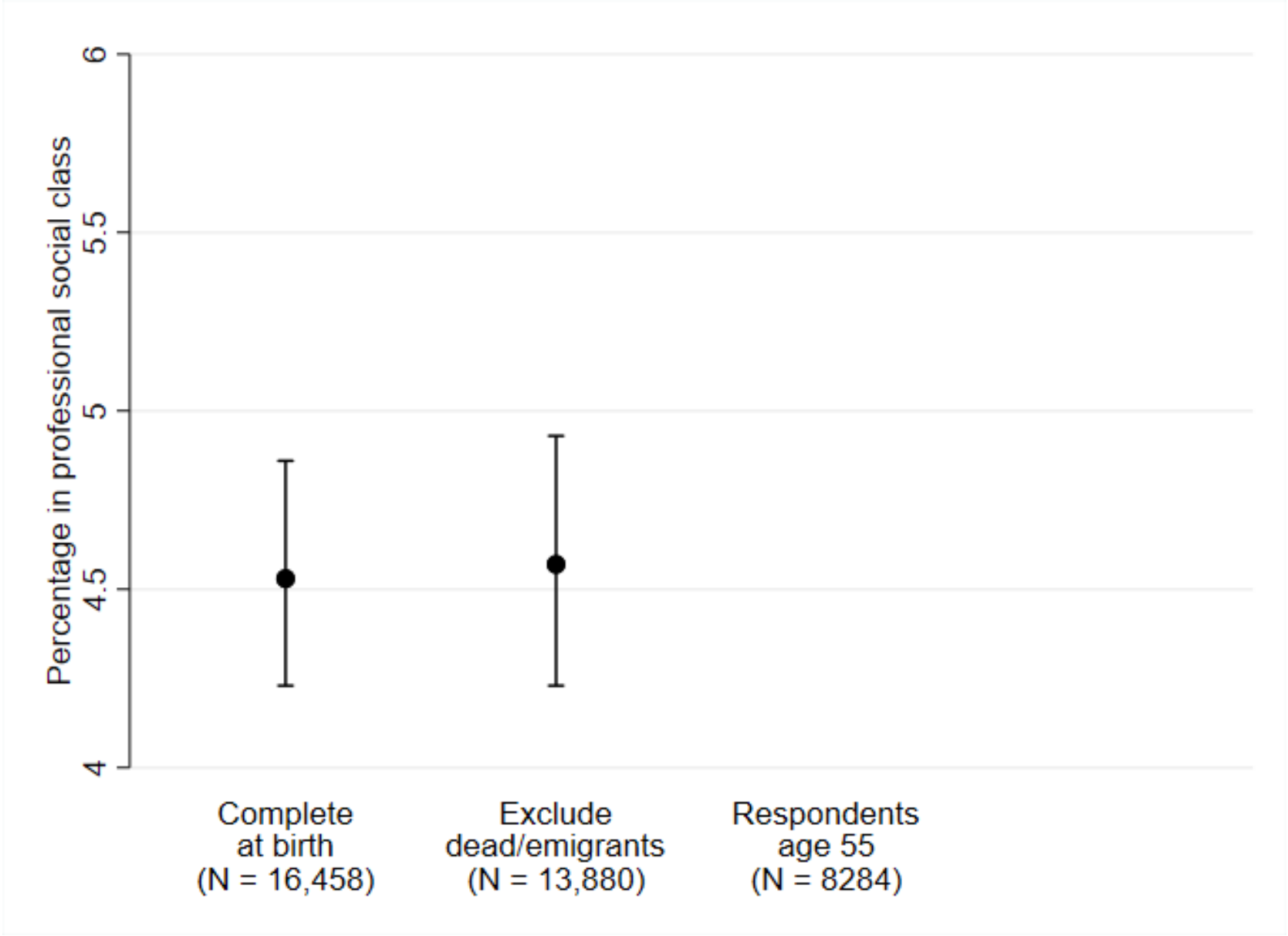
Social class of mother's husband at birth

CENTRE FOR LONGITUDINAL STUDIES



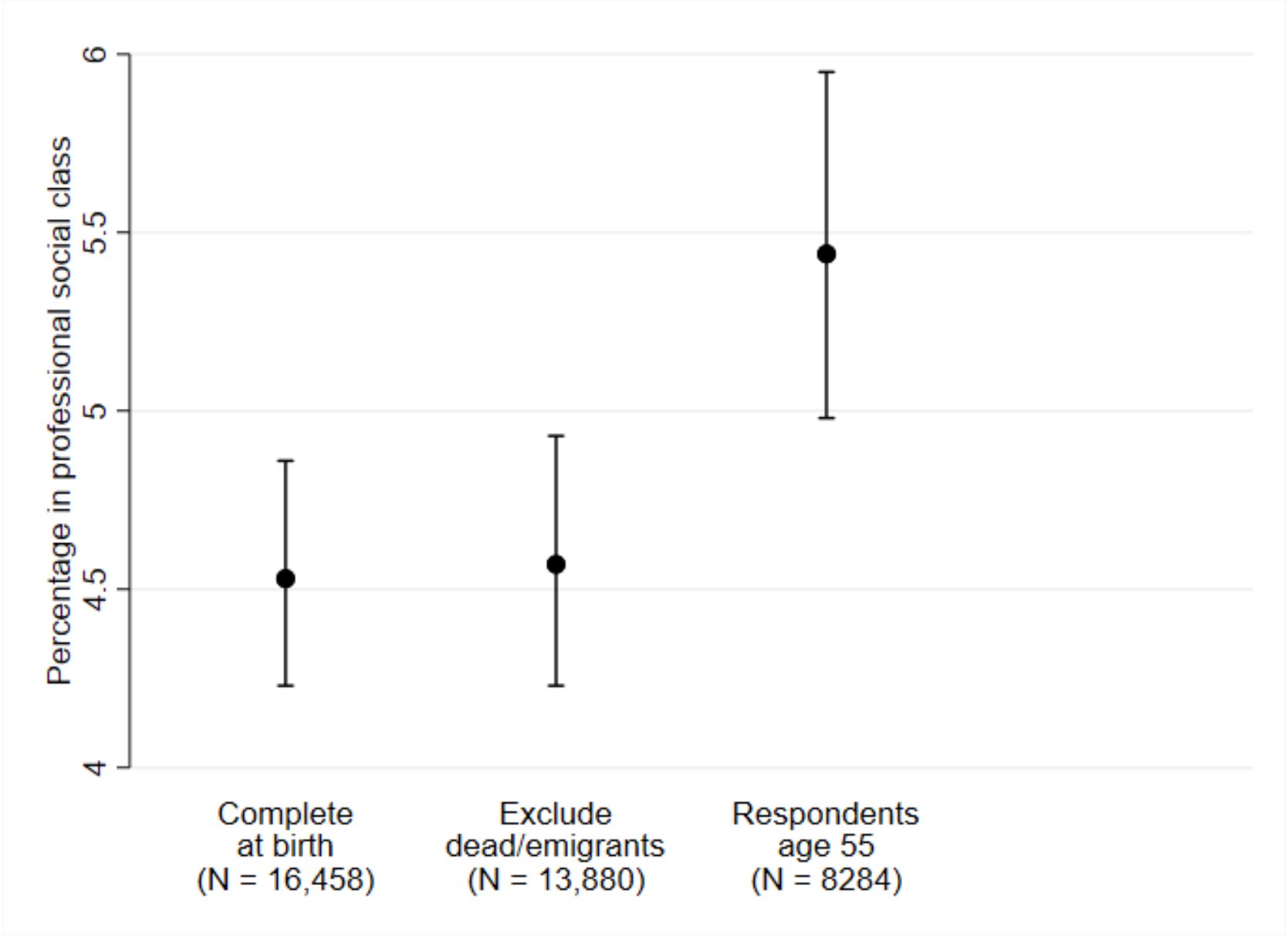
Social class of mother's husband at birth

CENTRE FOR LONGITUDINAL STUDIES



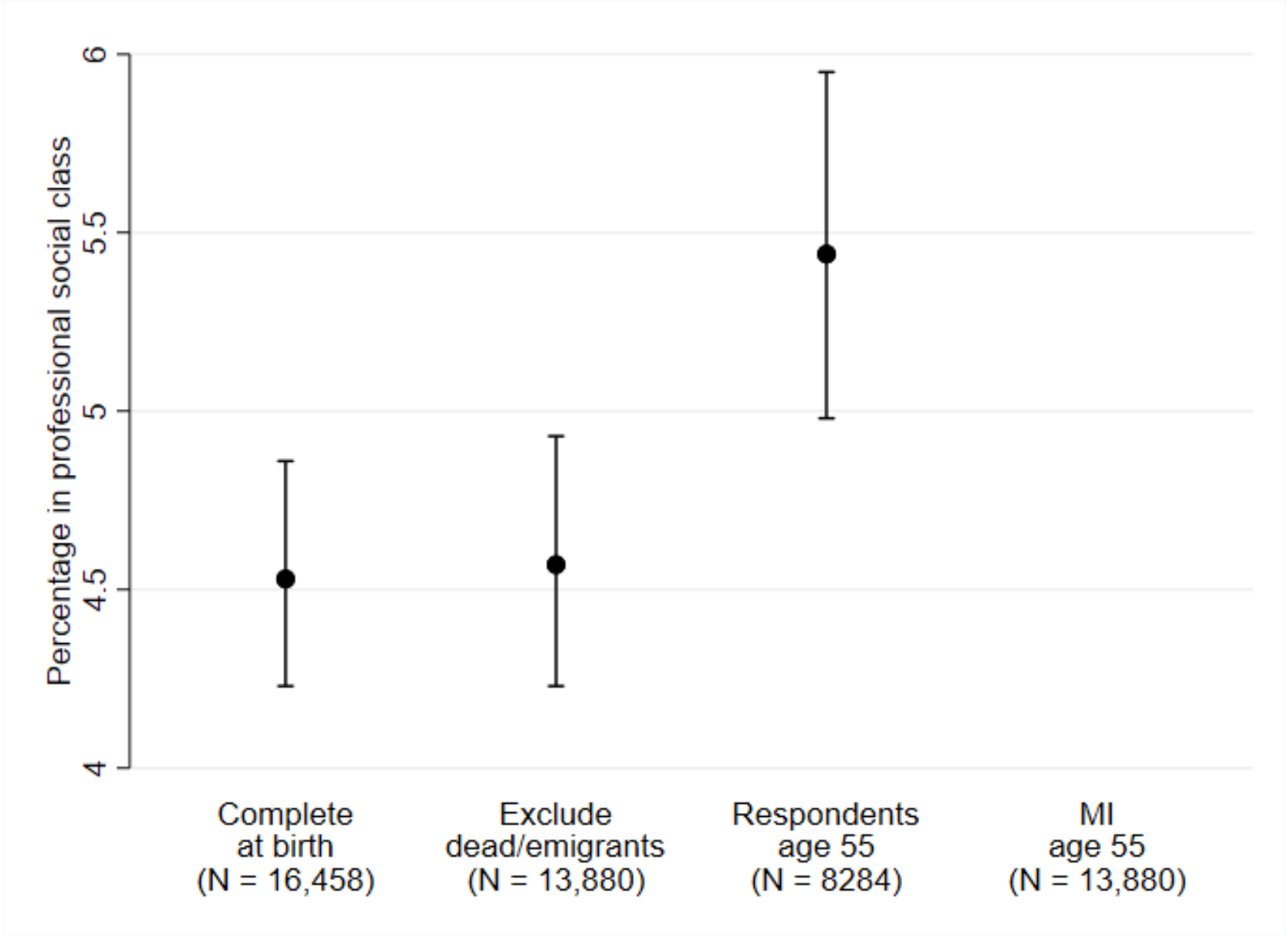
Social class of mother's husband at birth

CENTRE FOR LONGITUDINAL STUDIES



Social class of mother's husband at birth

CENTRE FOR LONGITUDINAL STUDIES



Multiple imputation

CENTRE FOR LONGITUDINAL STUDIES

- Complete case analysis only (generally) valid under MCAR.
- Many simple imputation approaches are problematic:
  - Mean imputation.
  - Last observation carried forward.
  - Single conditional imputation.
- In MI, plausible values are used in place of the missing values in a way that allows:
  1. Parameter estimates to be unbiased.
  2. Uncertainty to be estimated in an appropriate way.
- MI valid under MAR.

Multiple imputation

CENTRE FOR LONGITUDINAL STUDIES

- Specify an appropriate imputation model.
- Create a series of imputed datasets.
- Each imputed dataset analysed using the substantive model.
- Results combined using standard rules.
- Can be undertaken using standard statistical software.
- Widely adopted as practical for applied researchers in a wide range of settings.

Which variables should be included in the imputation model?

CENTRE FOR LONGITUDINAL STUDIES

Definitely:

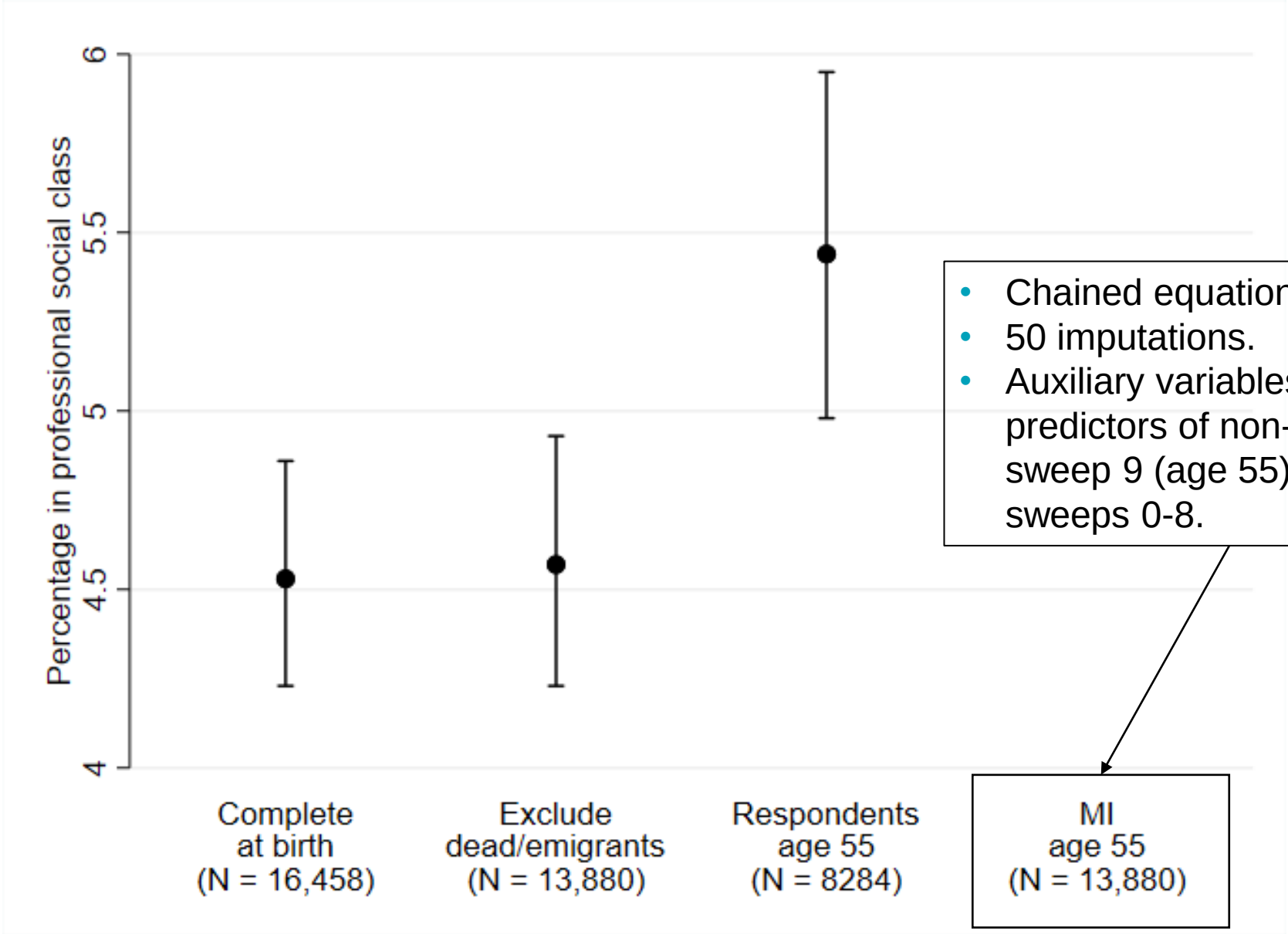
- All variables in the substantive model, including any interactions.

Optional “auxiliary variables”:

- Variables associated with the underlying values of the variable(s) subject to missingness.
- Particularly those also associated with the probability of missingness.

Social class of mother's husband at birth

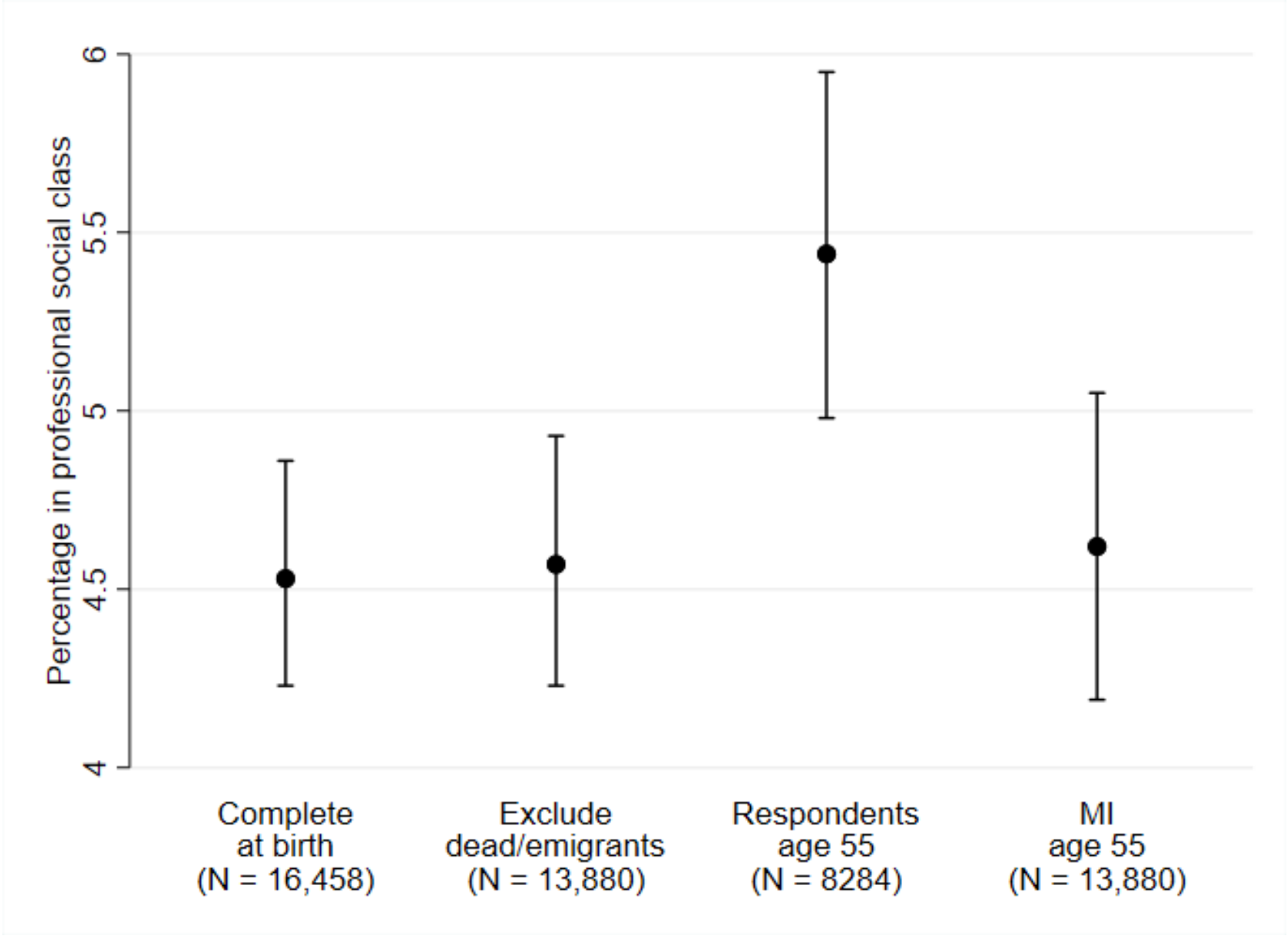
CENTRE FOR LONGITUDINAL STUDIES



- Chained equations.
- 50 imputations.
- Auxiliary variables: All predictors of non-response at sweep 9 (age 55) from sweeps 0-8.

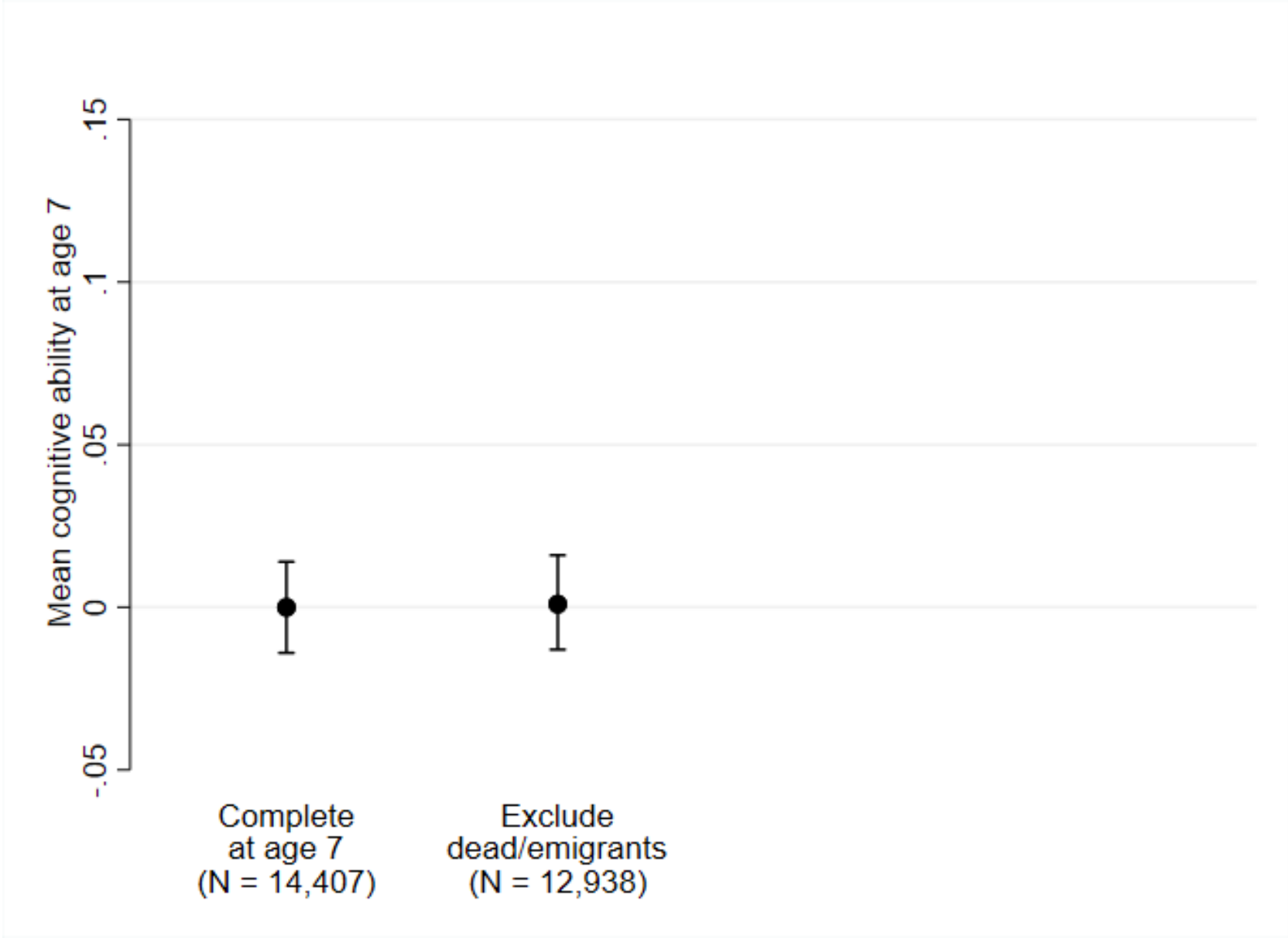
Social class of mother's husband at birth

CENTRE FOR LONGITUDINAL STUDIES



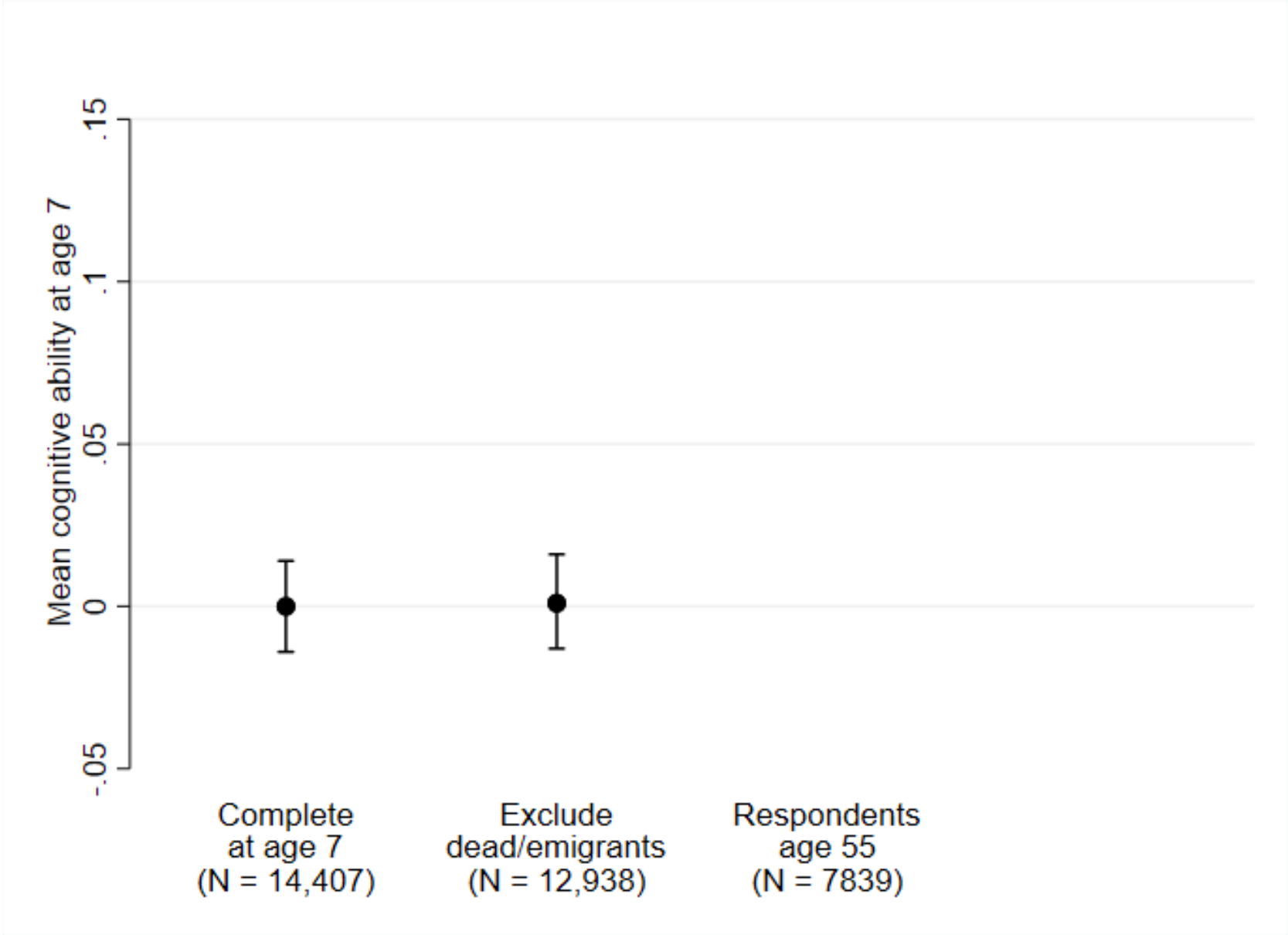
Cognitive ability at age 7

CENTRE FOR LONGITUDINAL STUDIES



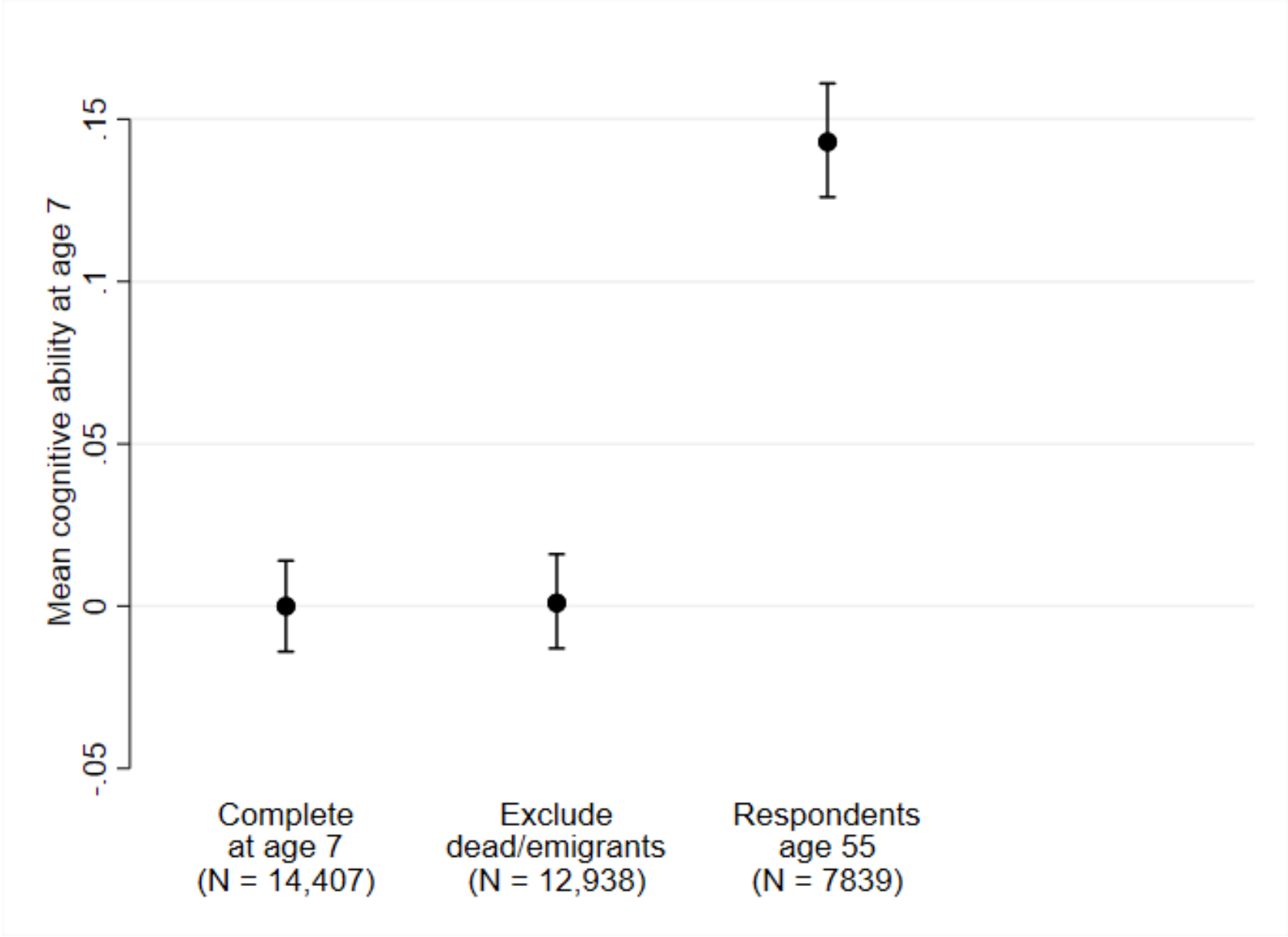
Cognitive ability at age 7

CENTRE FOR LONGITUDINAL STUDIES



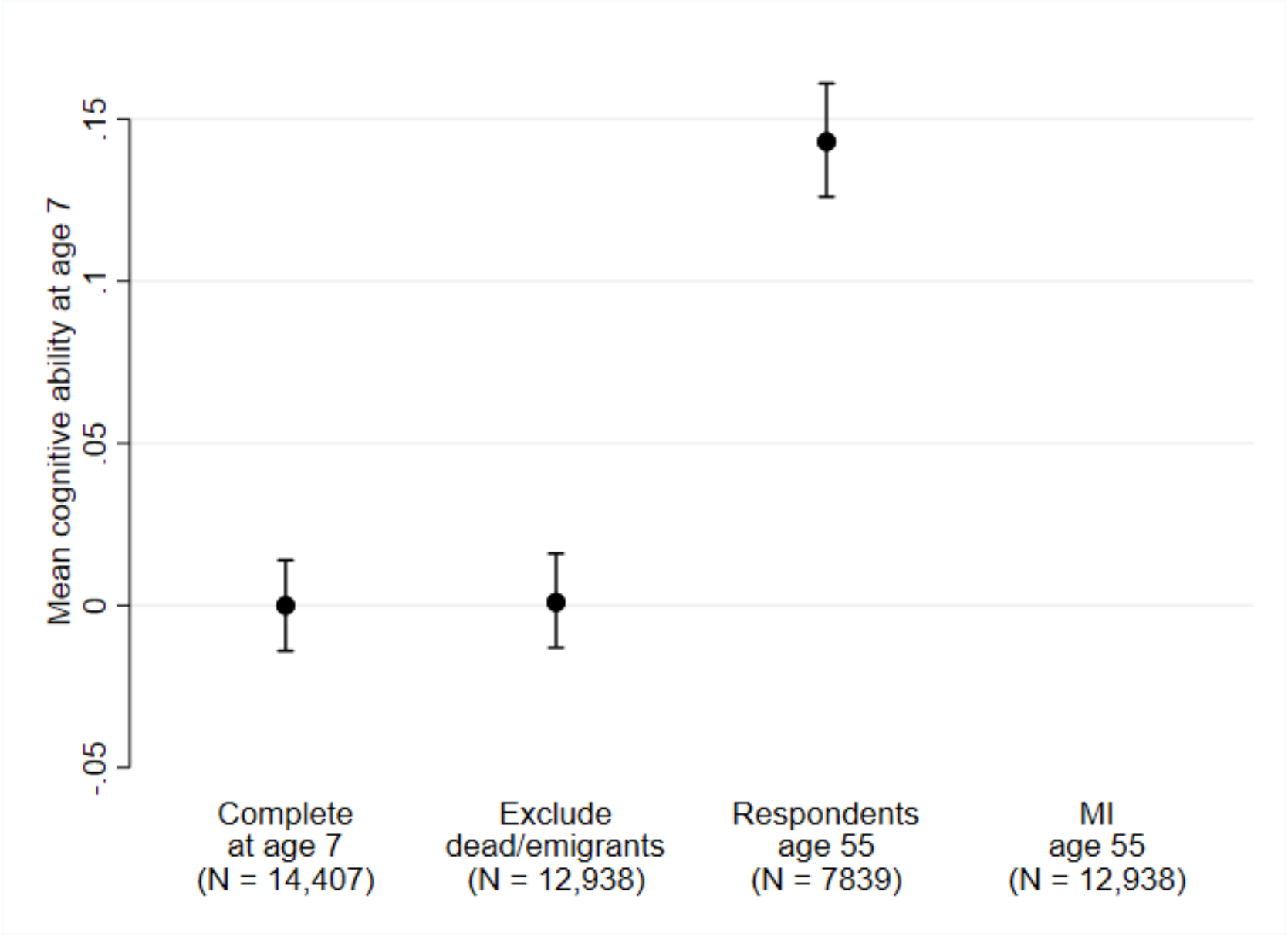
Cognitive ability at age 7

CENTRE FOR LONGITUDINAL STUDIES



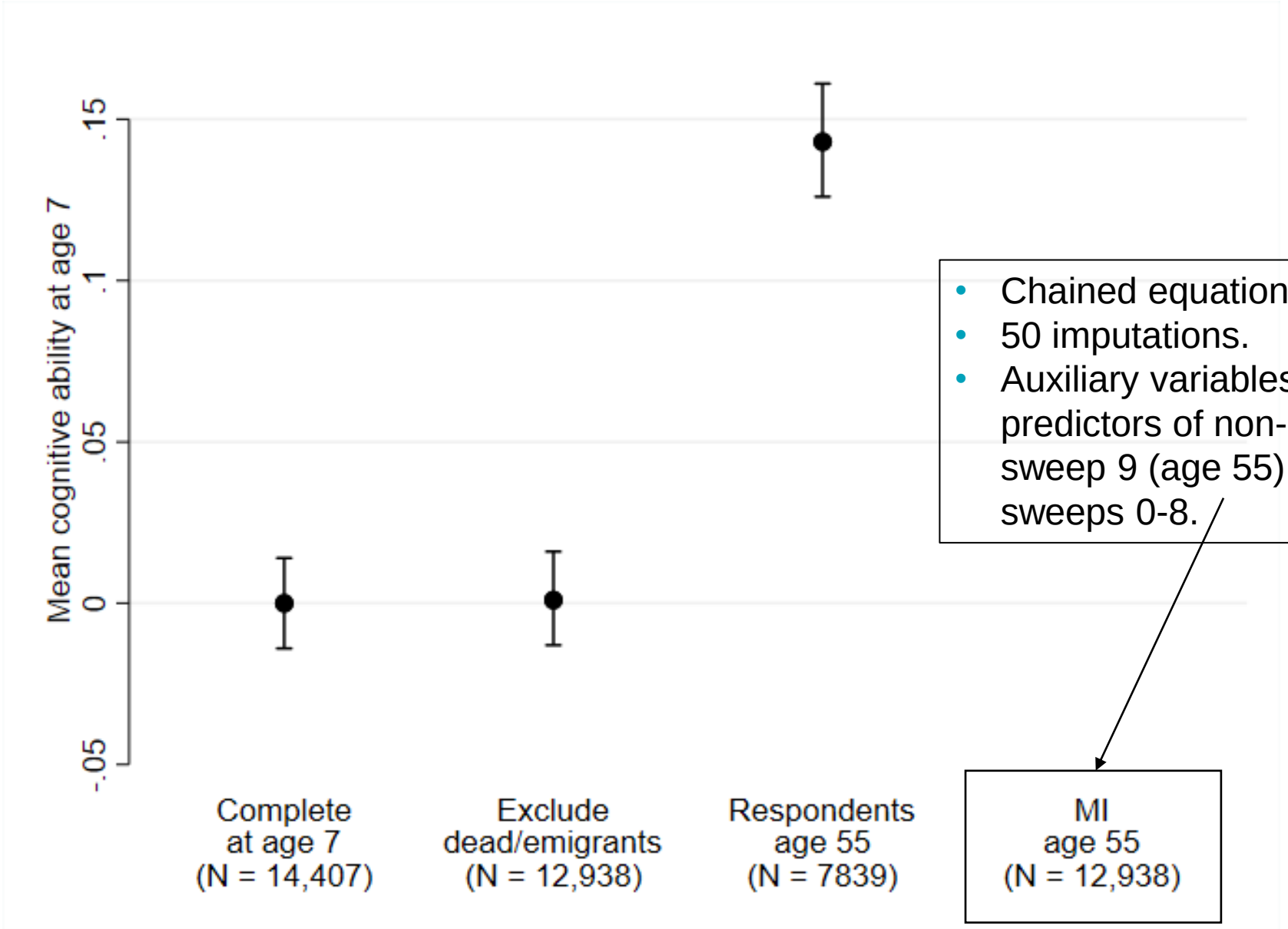
Cognitive ability at age 7

CENTRE FOR LONGITUDINAL STUDIES



Cognitive ability at age 7

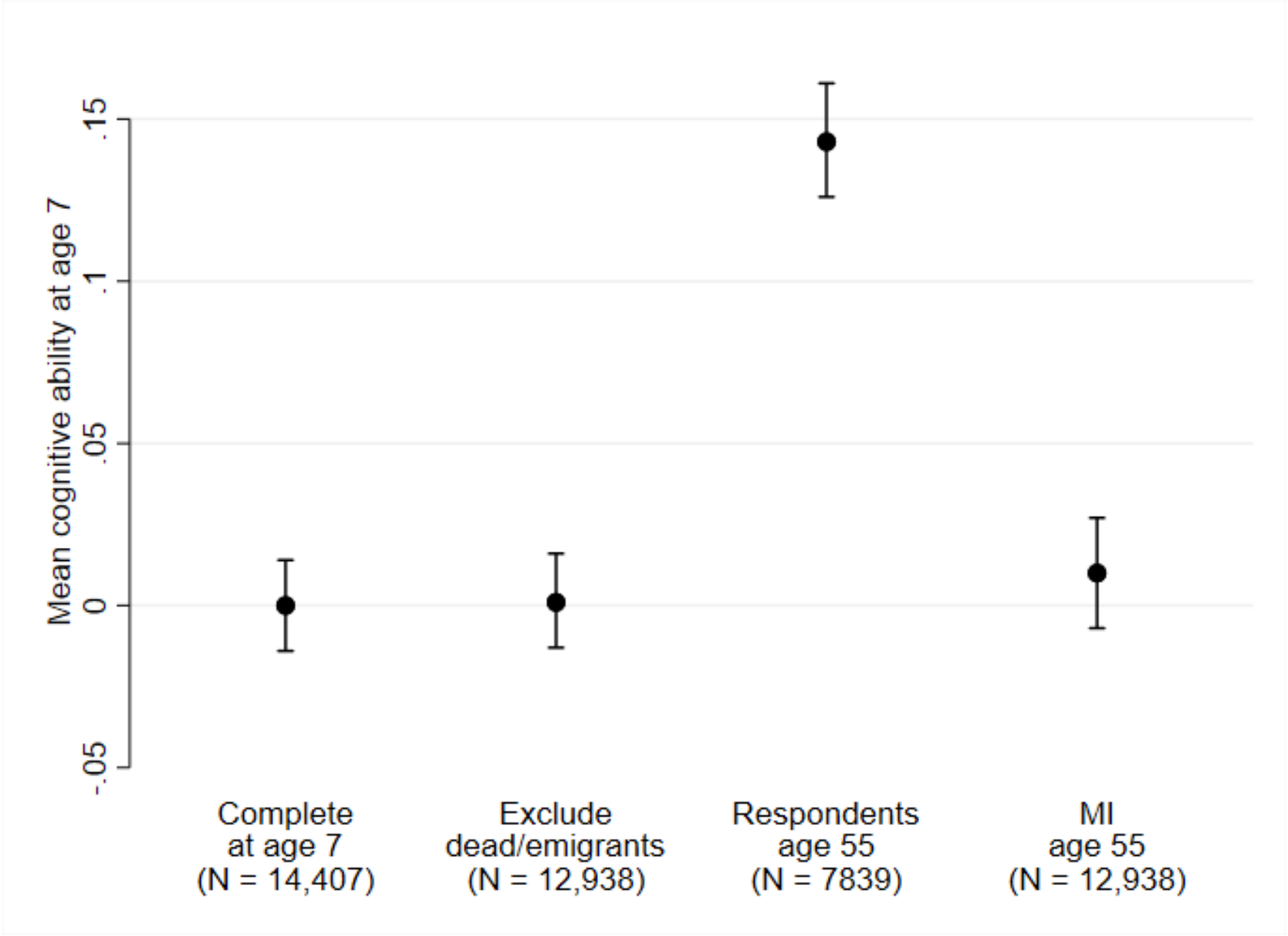
CENTRE FOR LONGITUDINAL STUDIES



- Chained equations.
- 50 imputations.
- Auxiliary variables: All predictors of non-response at sweep 9 (age 55) from sweeps 0-8.

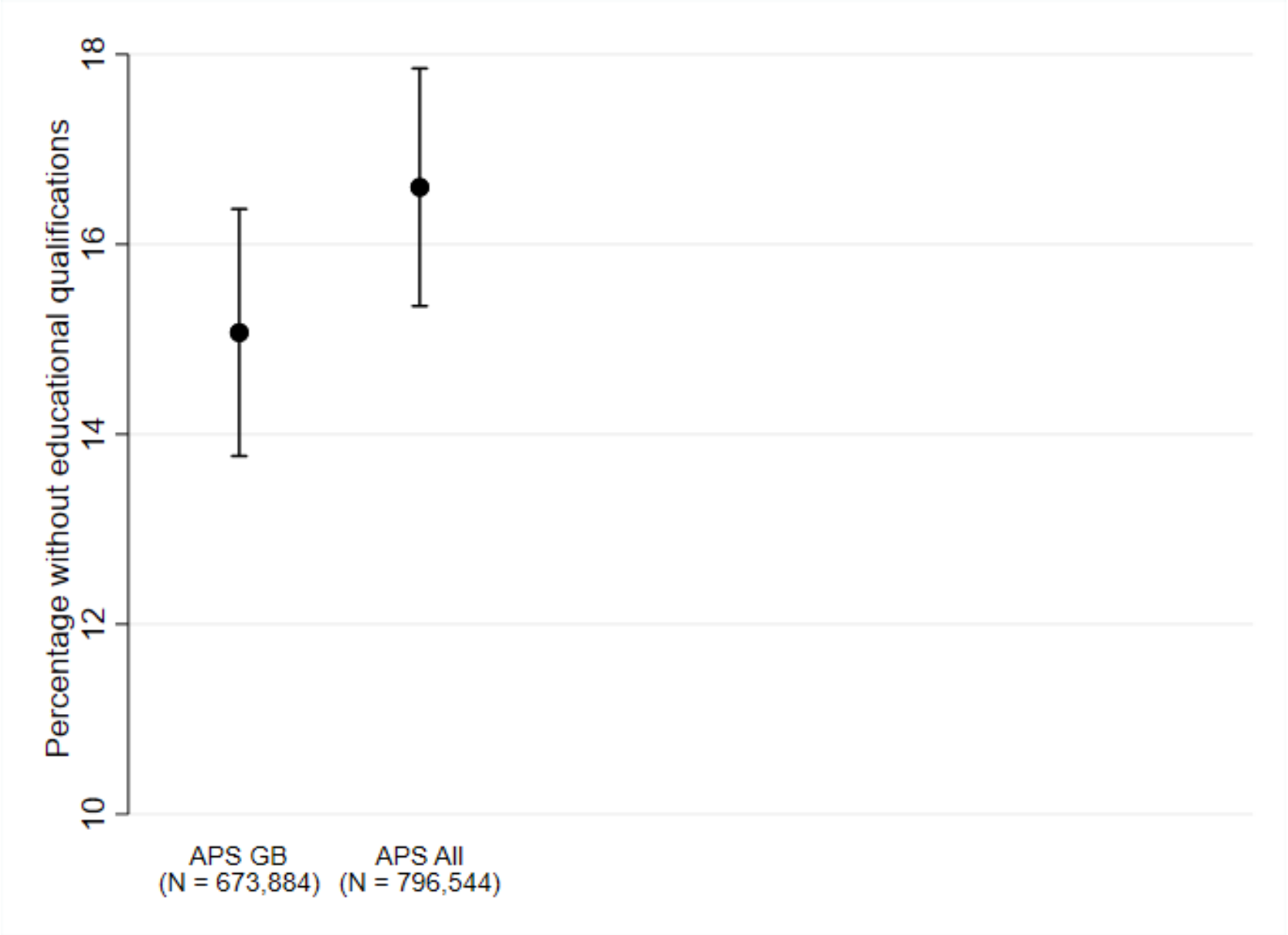
Cognitive ability at age 7

CENTRE FOR LONGITUDINAL STUDIES



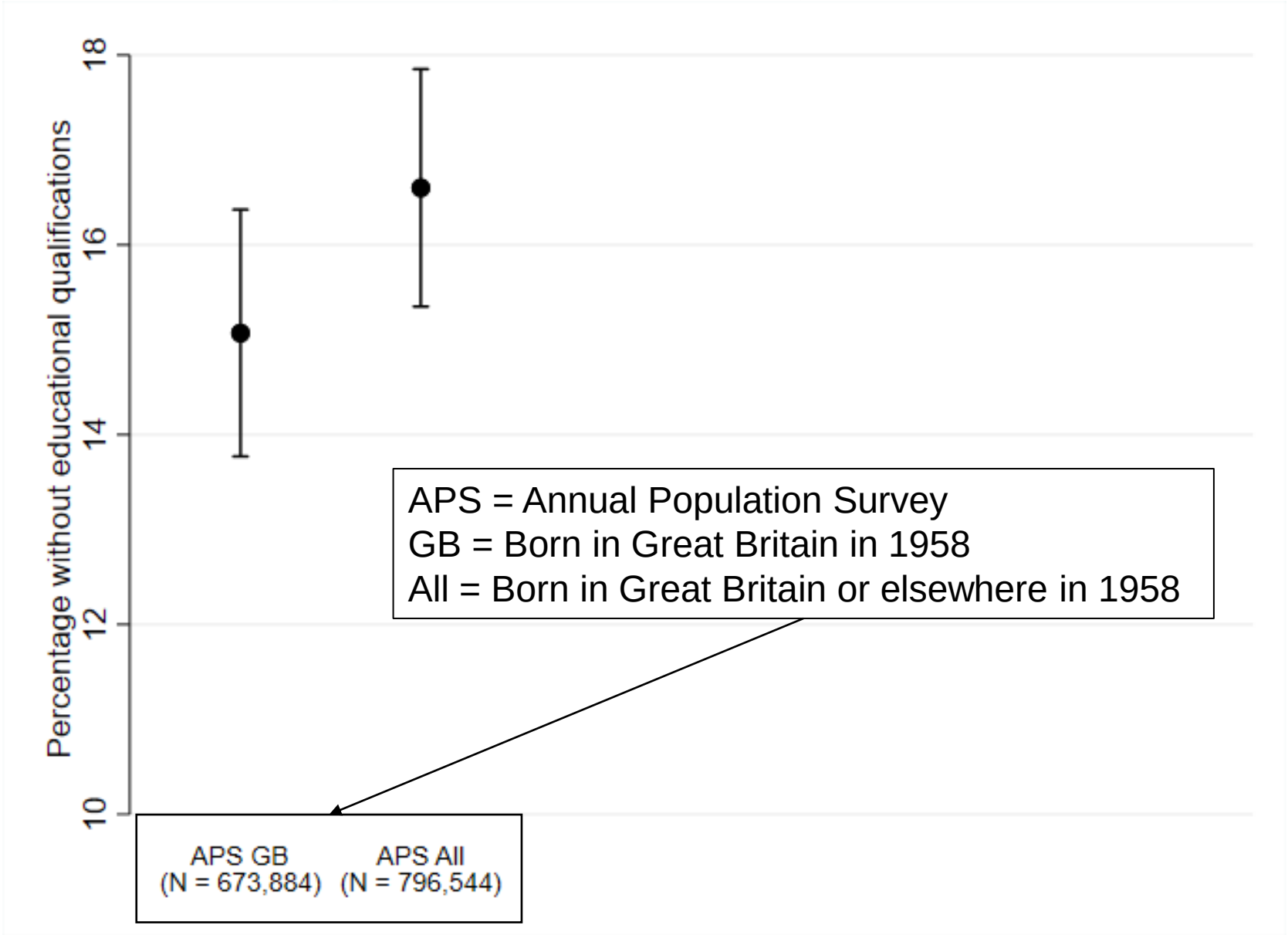
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



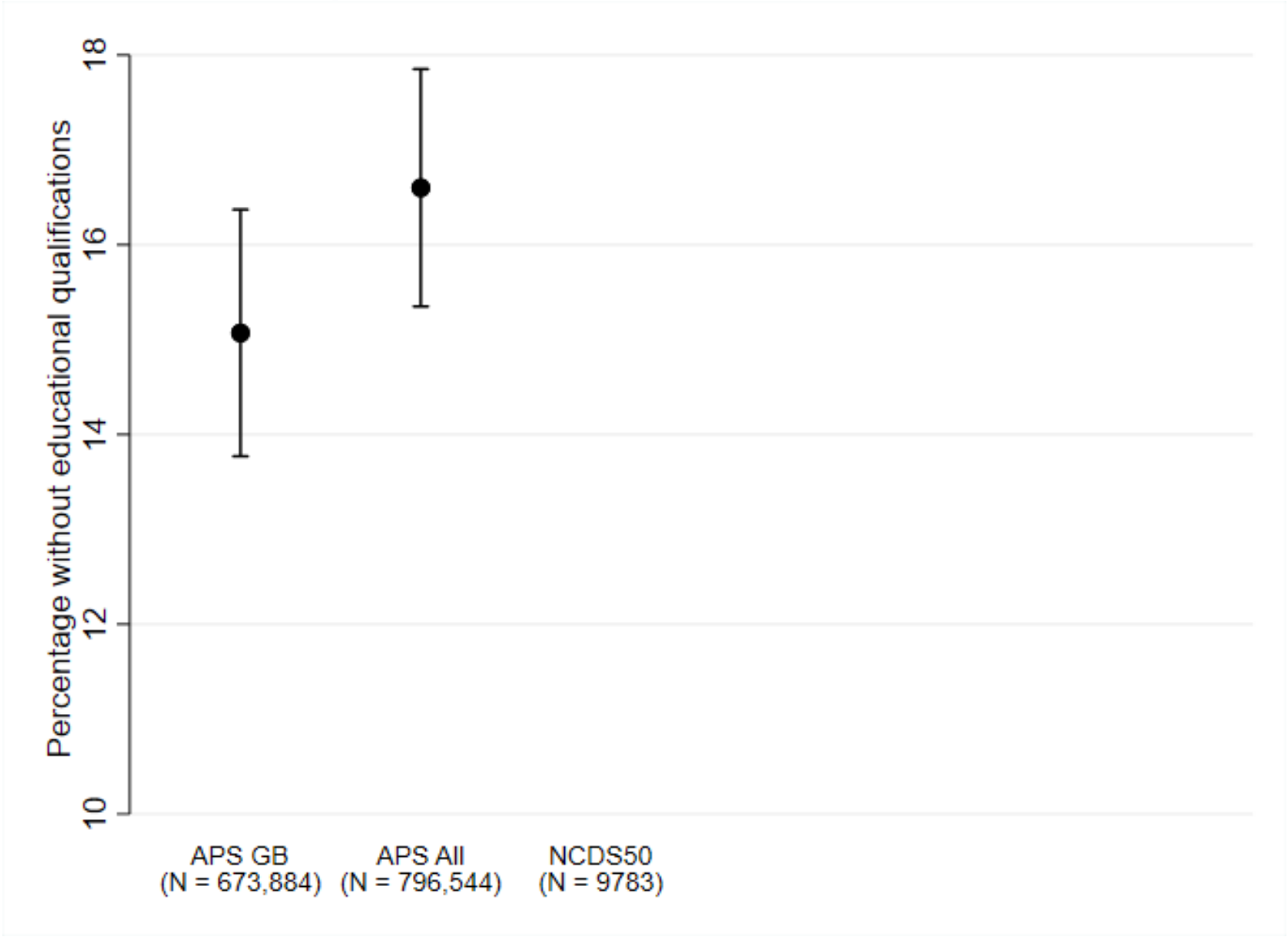
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



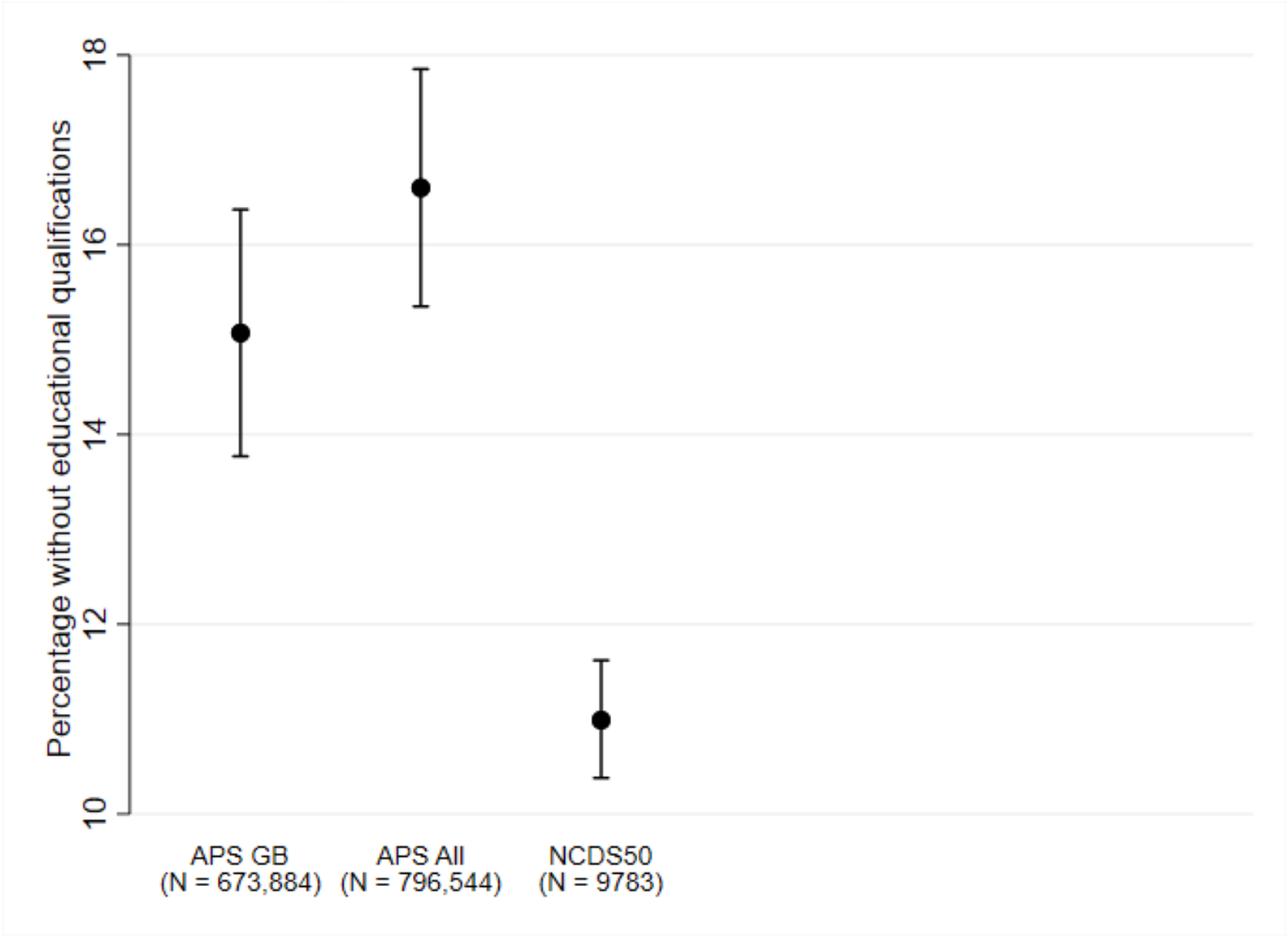
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



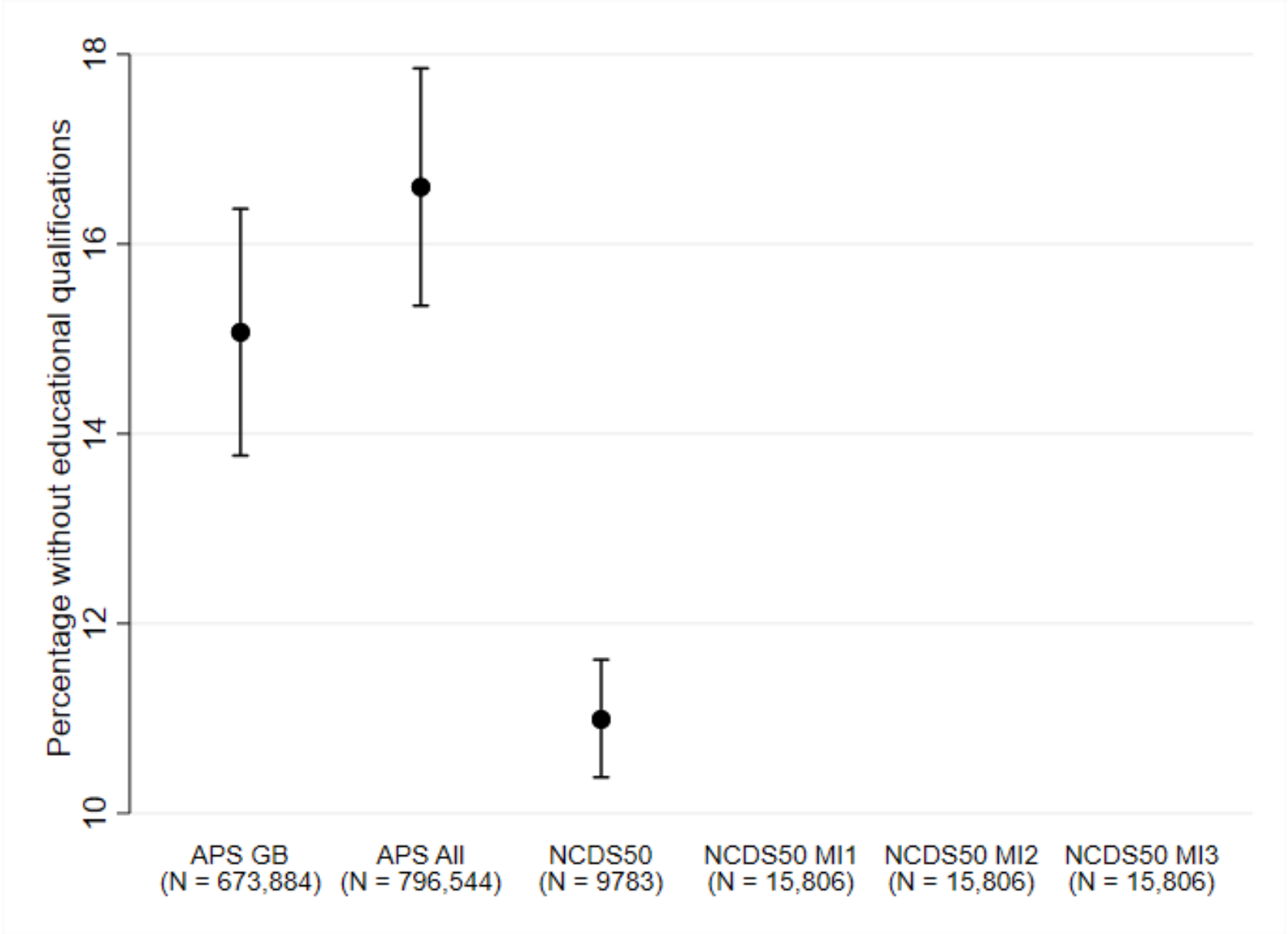
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



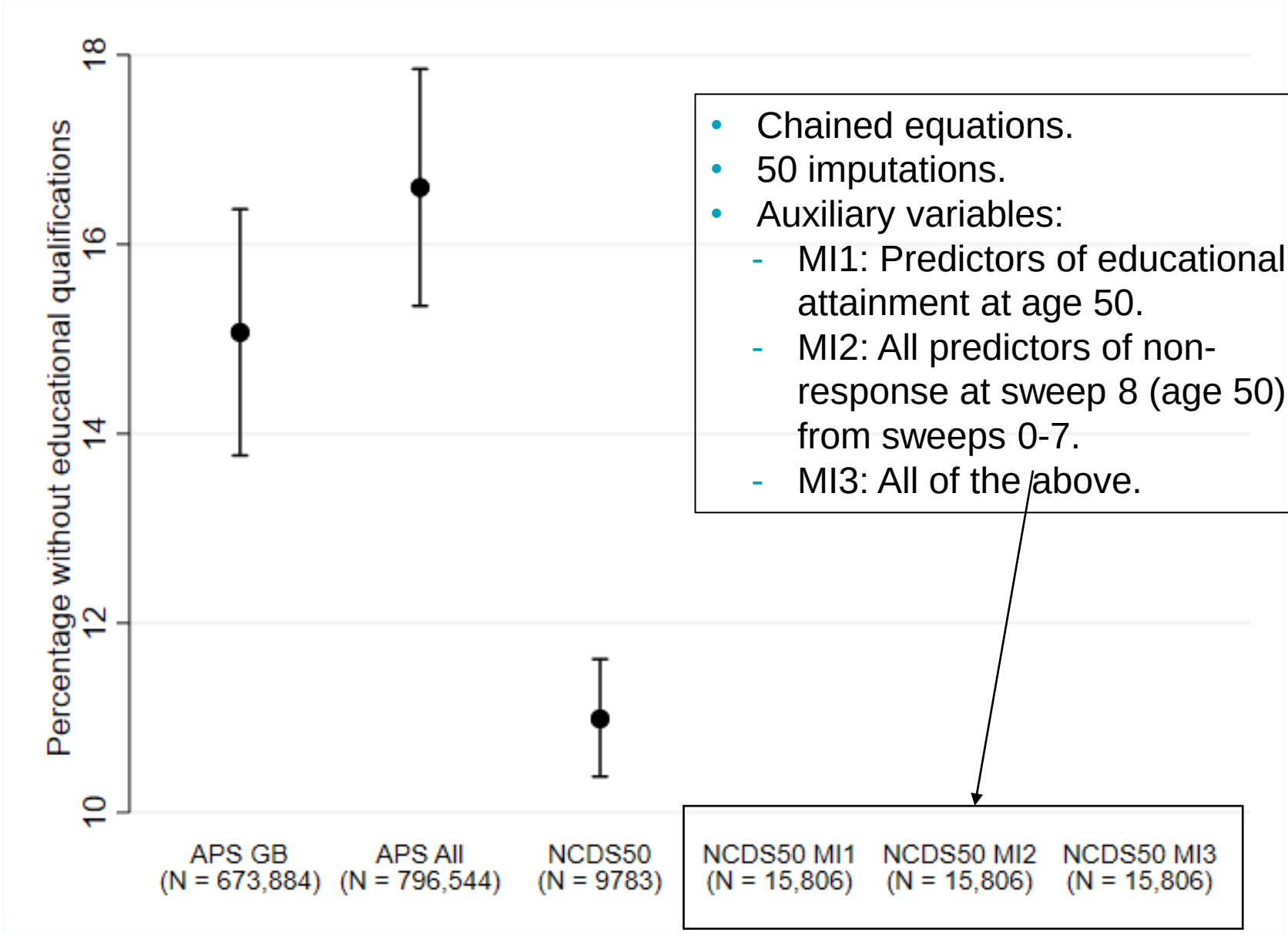
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



No educational qualifications at age 50

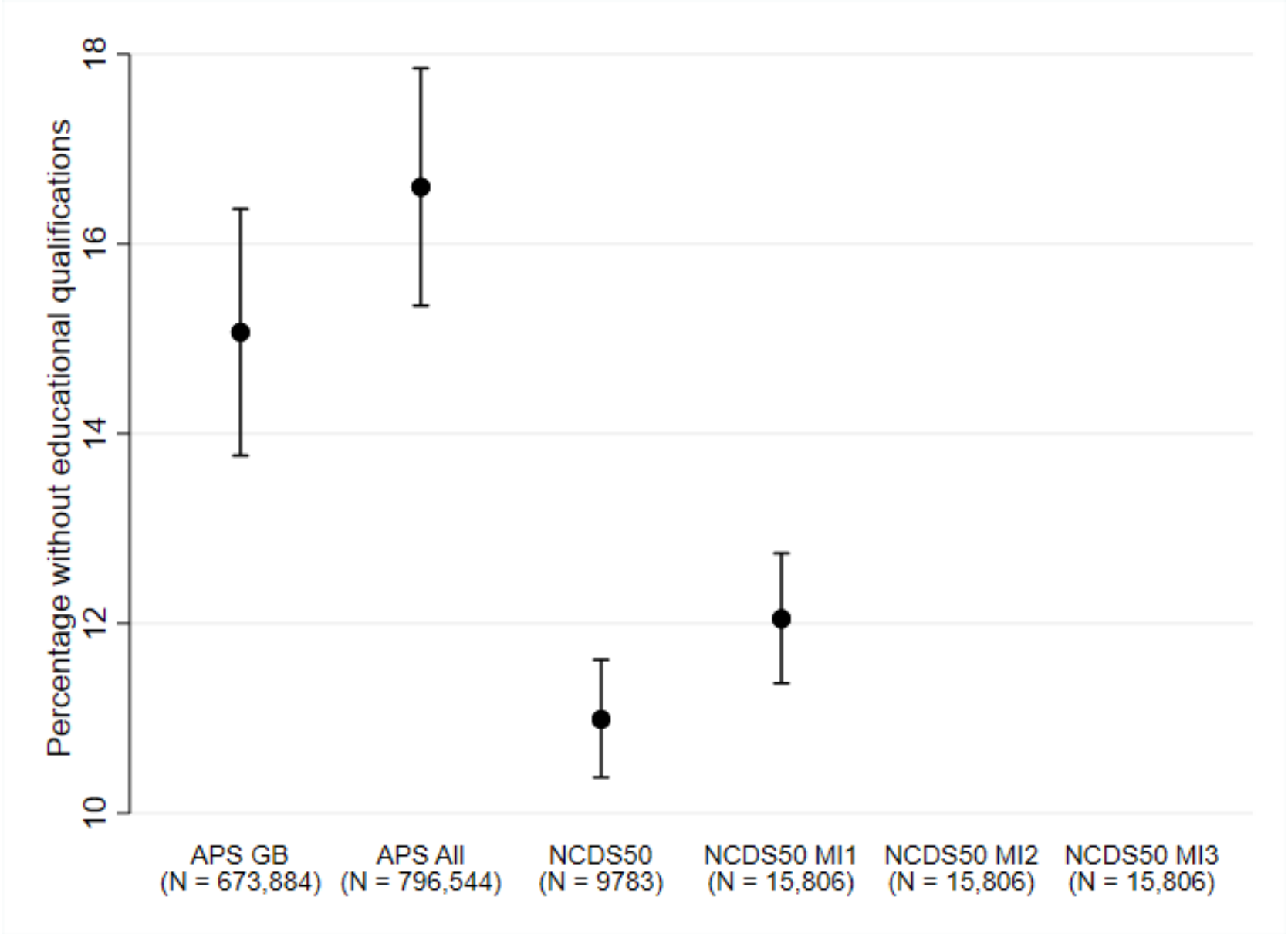
CENTRE FOR LONGITUDINAL STUDIES



- Chained equations.
- 50 imputations.
- Auxiliary variables:
  - MI1: Predictors of educational attainment at age 50.
  - MI2: All predictors of non-response at sweep 8 (age 50) from sweeps 0-7.
  - MI3: All of the above.

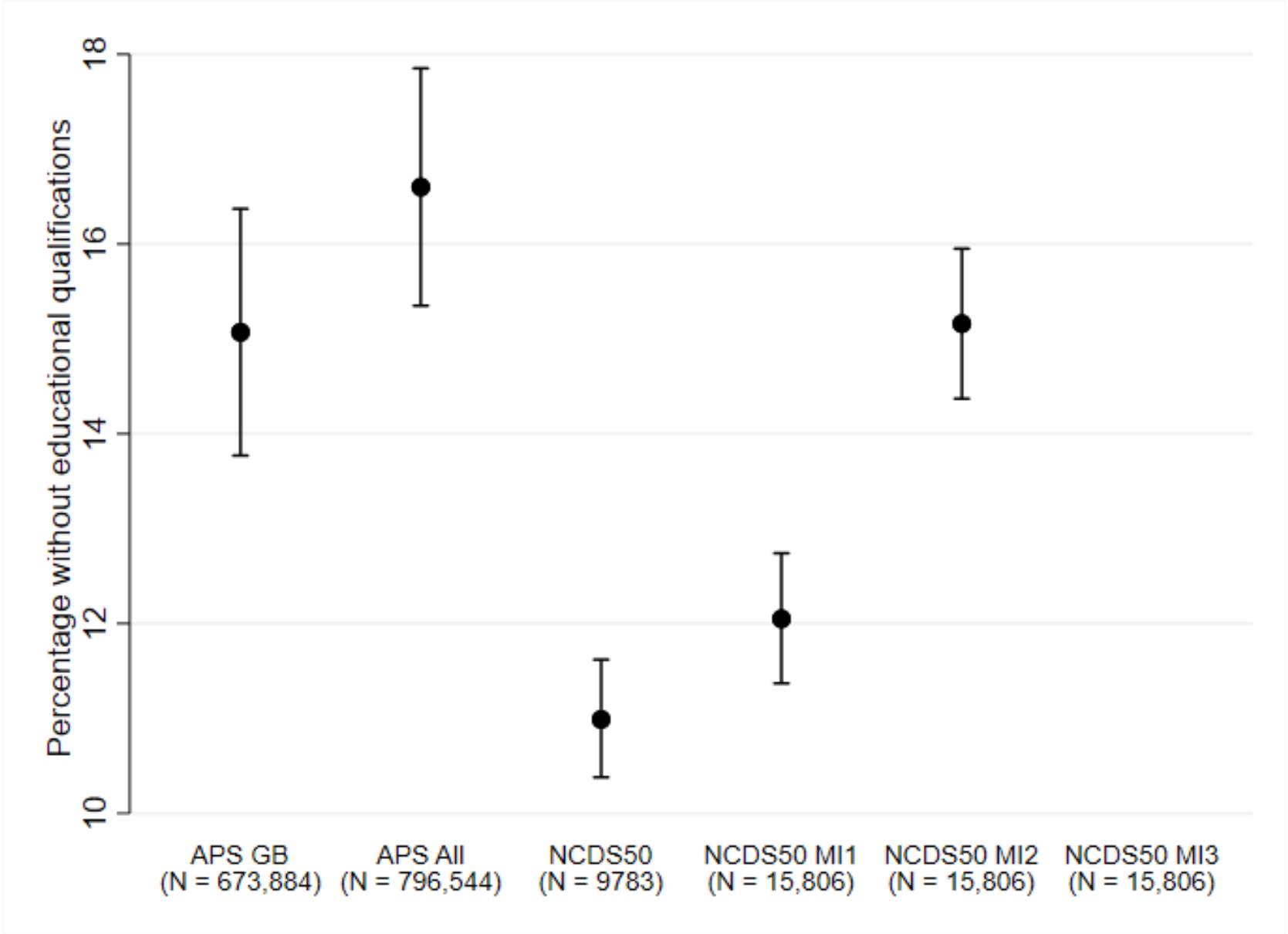
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



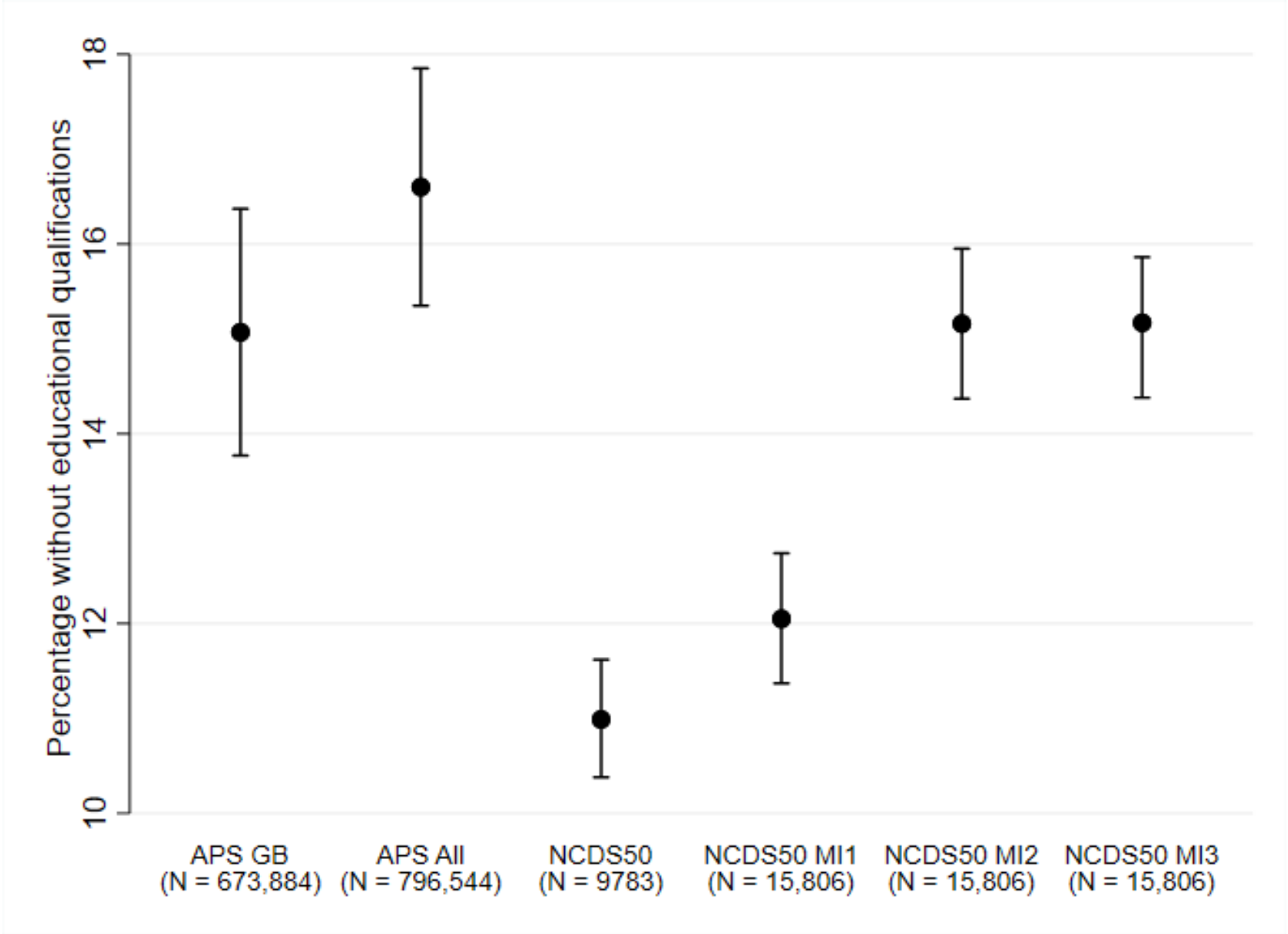
No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



No educational qualifications at age 50

CENTRE FOR LONGITUDINAL STUDIES



More realistic analyses

CENTRE FOR LONGITUDINAL STUDIES

- Such simple analyses are useful for testing/illustrating the basic idea.
- More realistic analyses likely to be more complicated:
  - More variables in substantive model – exposure(s), outcome(s), control variables,…
  - Inclusion of auxiliary variable(s) predictive of non-response at further sweeps.
  - Inclusion of auxiliary variable(s) predictive of the underlying values of the variable(s) subject to missingness.
  - Inclusion of auxiliary variable(s) to deal with item non-response.
  - Different types of substantive model.

More realistic analyses

CENTRE FOR LONGITUDINAL STUDIES

- Basic idea remains the same.
- Main concern likely to be (in the MI setting) instability of the imputation model caused by number/type of variables.
- Illustrative realistic example used throughout the NCDS Missing Data User Guide (next session).

Summary

CENTRE FOR LONGITUDINAL STUDIES

- We have identified variables which predict non-response at each sweep of NCDS.
- These can be used as auxiliary variables in subsequent analyses to increase the plausibility of the MAR assumption.
- Simple test analyses have shown this approach to perform well.
- A straightforward approach, easily implemented in standard software.
- Lists of predictors of non-response available via NCDS Missing Data User Guide.
- Will be updated when new sweeps of data become available.
- Also work using linked data (see later).

1958 National Child Development Study (NCDS)

CENTRE FOR LONGITUDINAL STUDIES

ELSEVIER

Journal of Clinical Epidemiology 136 (2021) 44–54

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Missing at random assumption made more plausible: evidence from the 1958 British birth cohort

Tarek Mostafa, Martina Narayanan, Benedetta Pongiglione, Brian Dodgeon, Alissa Goodman, Richard J. Silverwood*, George B. Ploubidis

University College London, London, UK

Accepted 24 February 2021; Available online 27 February 2021

Abstract

Objective: Non-response is unavoidable in longitudinal surveys. The consequences are lower statistical power and the potential for bias. We implemented a systematic data-driven approach to identify predictors of non-response in the National Child Development Study (NCDS; 1958 British birth cohort). Such variables can help make the missing at random assumption more plausible, which has implications for the handling of missing data

Study Design and Setting: We identified predictors of non-response using data from the 11 sweeps (birth to age 55) of the NCDS (n = 17,415), employing parametric regressions and the LASSO for variable selection.

Results: Disadvantaged socio-economic background in childhood, worse mental health and lower cognitive ability in early life, and lack of civic and social participation in adulthood were consistently associated with non-response. Using this information, along with other data from NCDS, we were able to replicate the "population distribution" of educational attainment and marital status (derived from external data), and the original distributions of key early life characteristics.

Conclusion: The identified predictors of non-response have the potential to improve the plausibility of the missing at random assumption. They can be straightforwardly used as "auxiliary variables" in analyses with principled methods to reduce bias due to missing data.

Keywords: Cohort studies; Longitudinal data; Missing data; Multiple imputation; National Child Development Study; Non-response

1. Introduction

Non-response is unavoidable in longitudinal surveys. The consequences are smaller samples due to attrition, lower statistical power and decreased representativeness compared to the originally intended target population. With some exceptions where complete case analysis is valid [1-3], in the majority of analyses of longitudinal data bias will occur if the implications of selection due to incompleteness are not formally addressed [4,5]. There is a broad interdisciplinary consensus that missing data should be dealt with using principled approaches and it has recently been argued that "complete-case analysis should be used with the same caution we ascribe to unadjusted estimates, as its validity relies on strong, often unrealistic assumptions" [6].

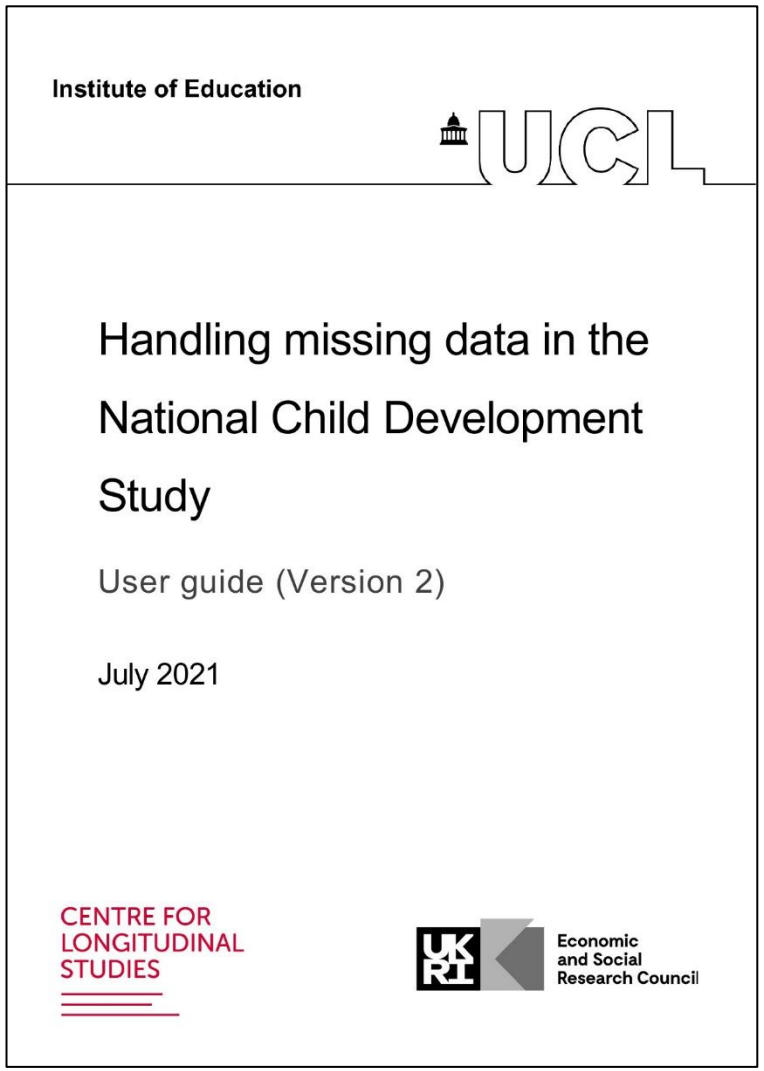
Rubin described 3 missing data generating mechanisms: i) missing completely at random (MCAR); ii) missing at random (MAR); iii) missing not at random (MNAR) [3,7,8]. MCAR implies that the probability of non-response does not depend on any variable (measured or unmeasured), or that there are no systematic differences between the observed and missing data. MCAR is partially testable, since we can examine whether variables available in our data are associated with missingness. MAR implies that systematic differences between the missing values and the observed values can be explained by observed data, or that given the observed data, the reasons for missingness do not depend on unobserved variables. With some exceptions for specific missing data patterns [9,10] the MAR assumption is untestable [11]. The third mechanism - MNAR - implies that that the observed data are insufficient to explain variation in the probability of missingness. MNAR is also untestable and methods to deal with this type of missing data generating mechanism rely heavily on further – usually distributional - assumptions [12].

Contextualizing the 1958 British National Child Development Study (NCDS) within Rubin's framework, we know that the missing data generating mechanism is not

Conflict of interest: None.

* Corresponding author.
E-mail address: R.Silverwood@ucl.ac.uk (R.J. Silverwood).

https://doi.org/10.1016/j.jclinepi.2021.02.019
0895-4356/



Institute of Education

UCL

Handling missing data in the National Child Development Study

User guide (Version 2)

July 2021

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

CENTRE FOR LONGITUDINAL STUDIES

1970 British Cohort Study (BCS70)

Identifying predictors of non-response in BCS

CENTRE FOR LONGITUDINAL STUDIES

- Very similar approach used in BCS as in NCDS
- Aim to maximise the plausibility of the MAR assumption using a data driven approach we identify the variables that are associated with non-response at each sweep (and can potentially be used as auxiliary variables)
- We also highlight cases in which we can explore MNAR

Identifying predictors of non-response in NCDS

CENTRE FOR LONGITUDINAL STUDIES

- ~20000 variables in BCS sweeps 0-8.
- Exclude:
  - Routed variables.
  - Binary variables with prevalence <1%.
  - Variables with item non-response > 40%.
- Use summary scores for scales
- For non-response at sweep *t* we used the same 3 stage approach as in NCDS (using a bit stricter criteria)

Predictors of non-response

CENTRE FOR LONGITUDINAL STUDIES

	NR sweep 1 (age 7)	NR sweep 2 (age 11)	NR sweep 3 (age 16)	NR sweep 4 (age 23)	NR sweep 5 (age 33)	NR sweep 6 (age 42)	NR sweep 7 (age 44)	NR sweep 8 (age 46)	NR sweep 9 (age 50)
Sweep 0 (birth)	4	3	1	5	1	5	4	2	5
Sweep 1 (age 7)		4	1	3	2	2	3	1	0
Sweep 2 (age 11)			1	3	1	1	3	1	1
Sweep 3 (age 16)				0	1	0	0	0	0
Sweep 4 (age 23)					2	3	2	1	0
Sweep 5 (age 33)						2	0	2	2
Sweep 6 (age 42)							0	0	2
Sweep 7 (age 46)								1	2
Sweep 8 (age 46)									1
Total	**4**	**7**	**3**	**11**	**7**	**13**	**12**	**8**	**13**

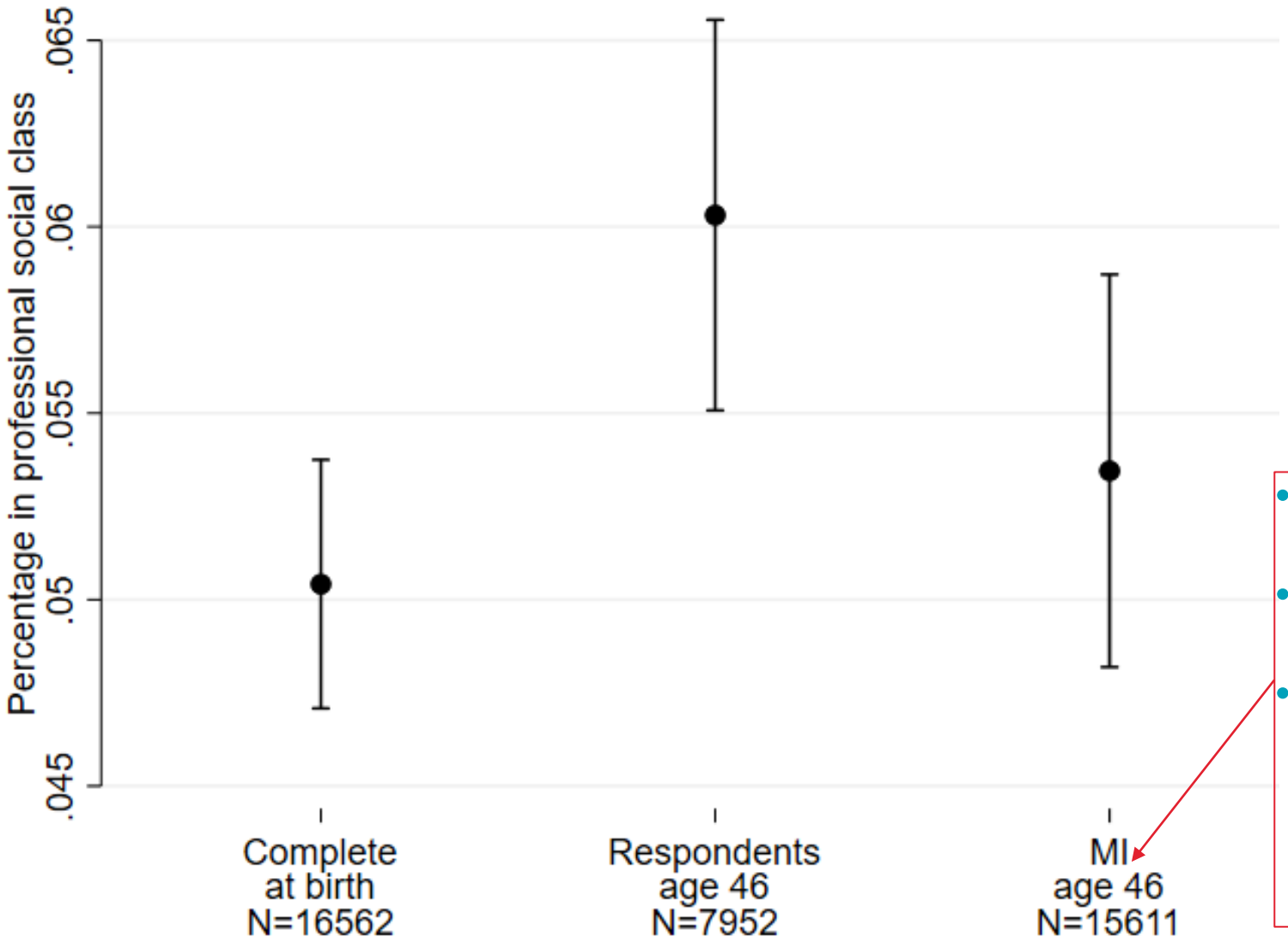
Consistent predictors of participation

CENTRE FOR LONGITUDINAL STUDIES

- Being female (adulthood only).
- Few household moves
- Paternal social class (early sweeps)
- Higher early life cognitive ability;
- Social participation - voting;
- Home ownership.
- Participation in previous sweeps.

Internal validation: Paternal Social class - at birth

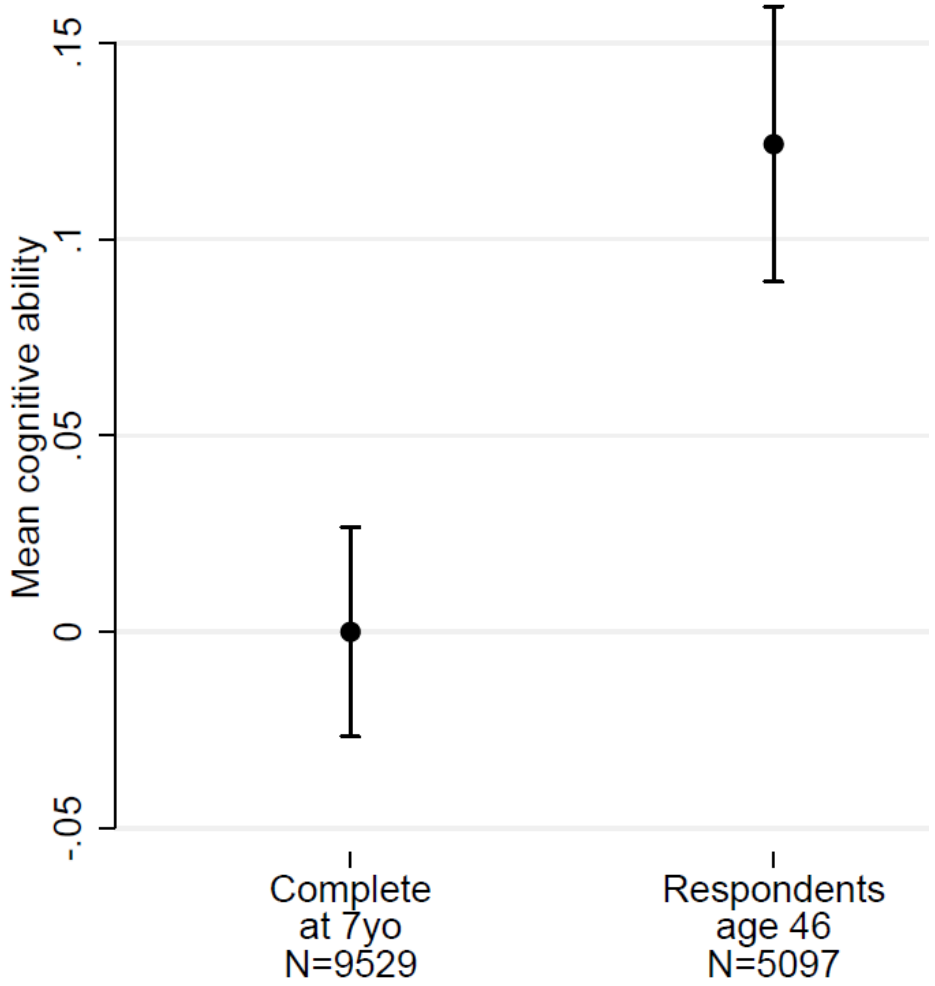
CENTRE FOR LONGITUDINAL STUDIES



- Chained equations.
- 50 imputations.
- Auxiliary variables: All predictors of non-response at sweep 9 (age 46) from sweeps 0-8.

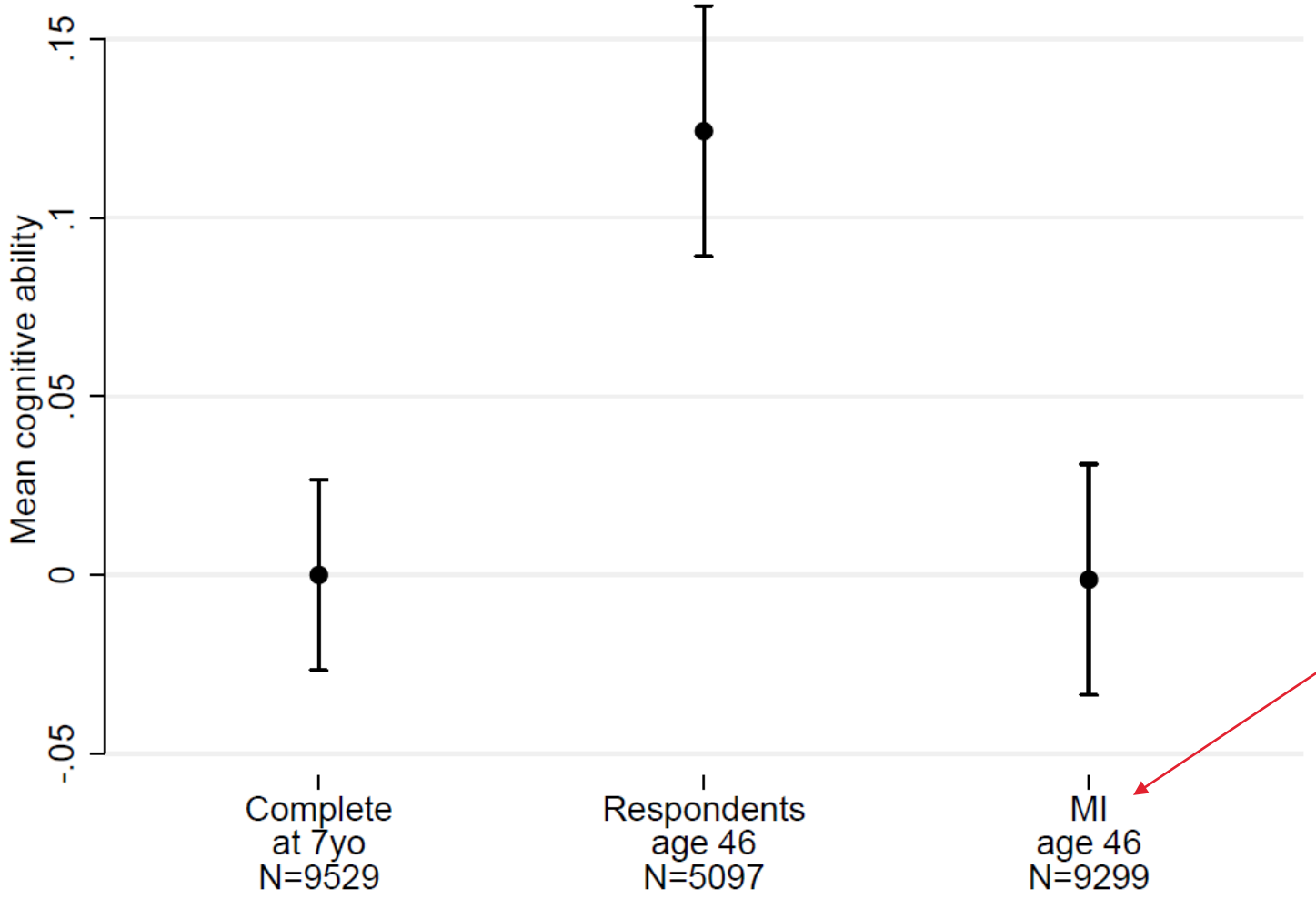
Internal validation: Cognitive ability - at 7yo

CENTRE FOR LONGITUDINAL STUDIES



Internal validation: Cognitive ability - at 7yo

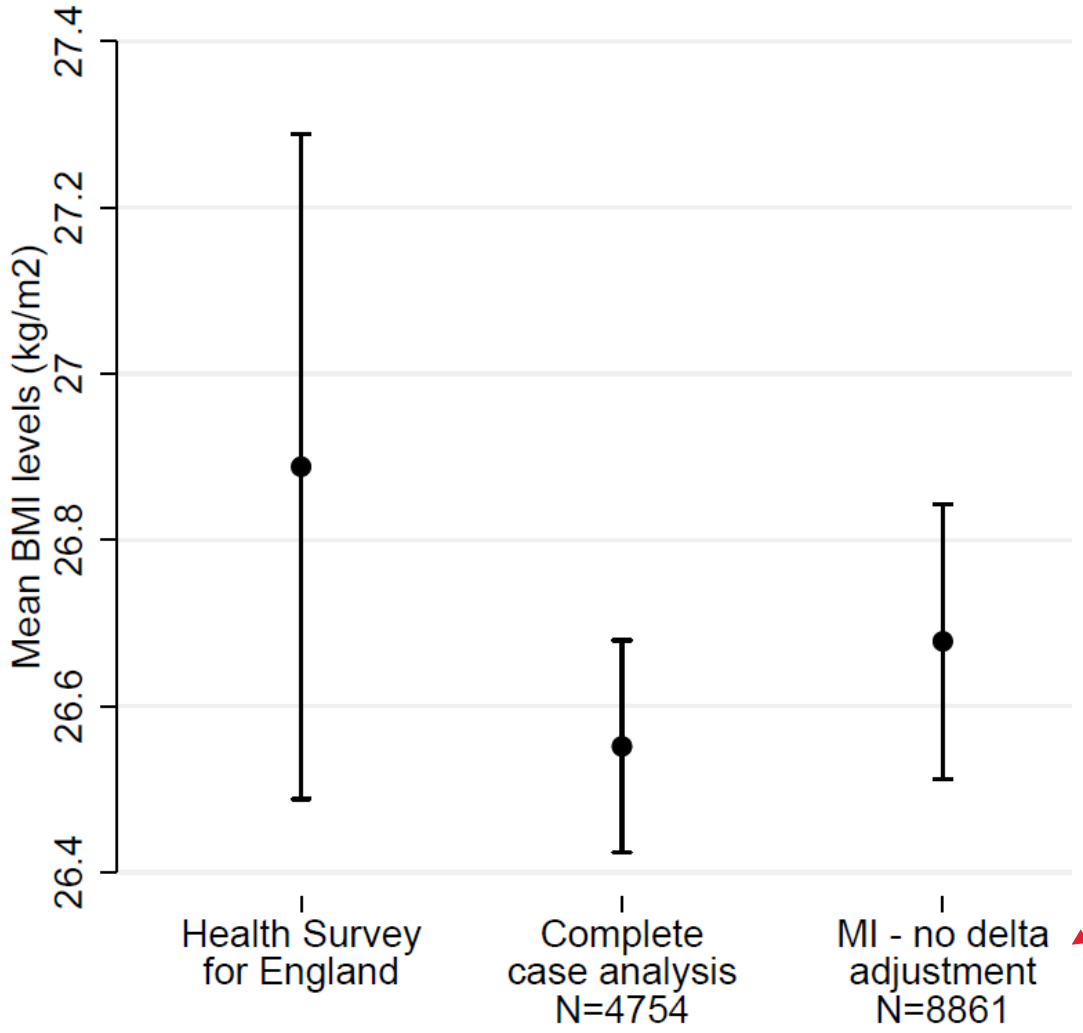
CENTRE FOR LONGITUDINAL STUDIES



- Chained equations.
- 50 imputations.
- Auxiliary variables: All predictors of non-response at sweep 9 (age 46) from sweeps 0-8.

Mean BMI levels – Age 34 (MEN)

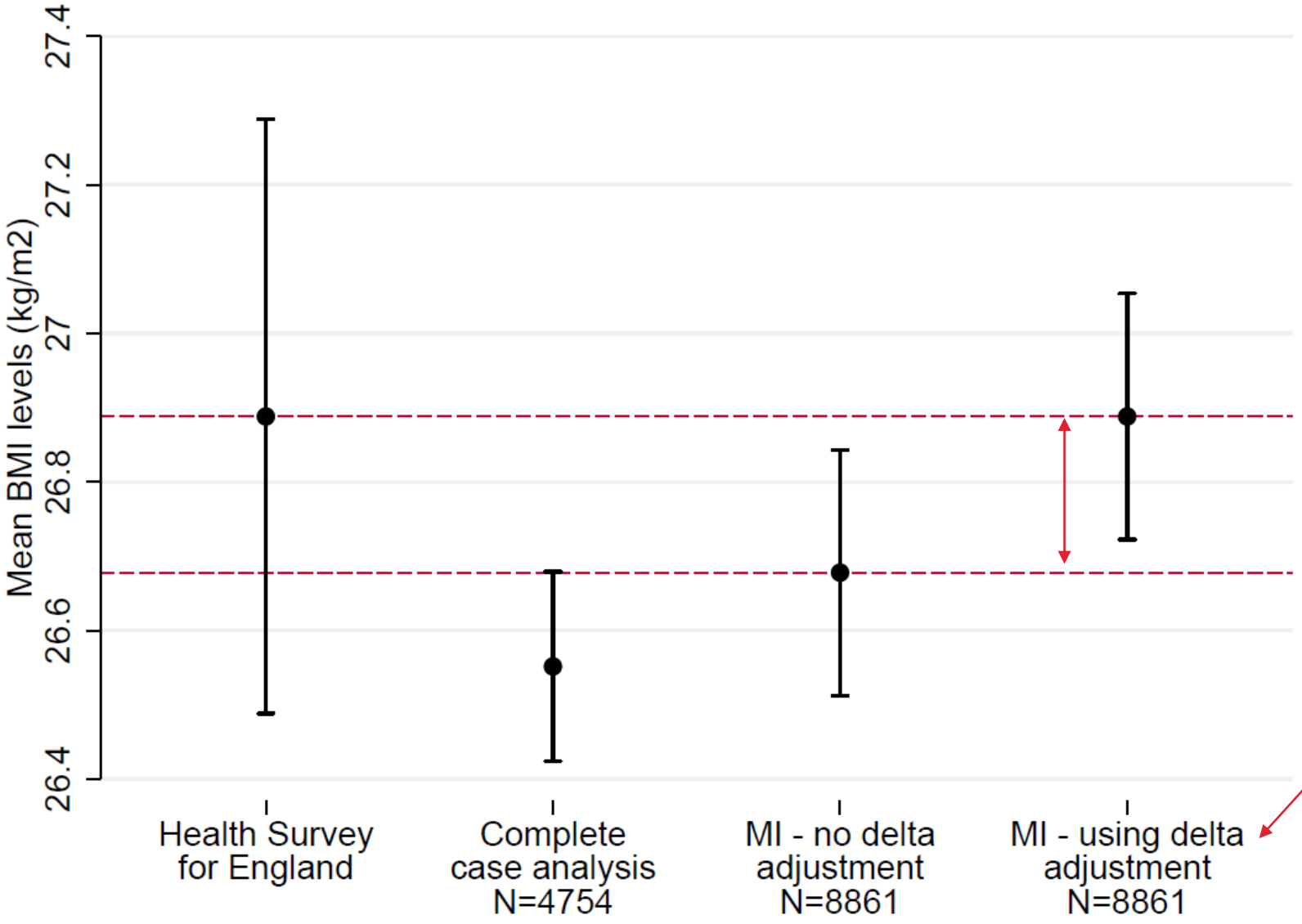
CENTRE FOR LONGITUDINAL STUDIES



- Chained equations.
- 50 imputations.
- Auxiliary variables: All predictors of non-response at sweep 6 (age 34) from sweeps 0-5.

SENSITIVITY ANALYSIS – External validation
BMI levels – Age 34 (MEN)

CENTRE FOR LONGITUDINAL STUDIES



We added a delta value (~0.4kg/m2) in the imputed data in the obs with missing data, after MI, so that MI using delta adj results match the results from HSE

SENSITIVITY ANALYSIS – External validation: Mean BMI levels

CENTRE FOR LONGITUDINAL STUDIES

We followed the same procedure for

- Women at age 34
- Men at age 42
- Women at age 42

Summary

CENTRE FOR LONGITUDINAL STUDIES

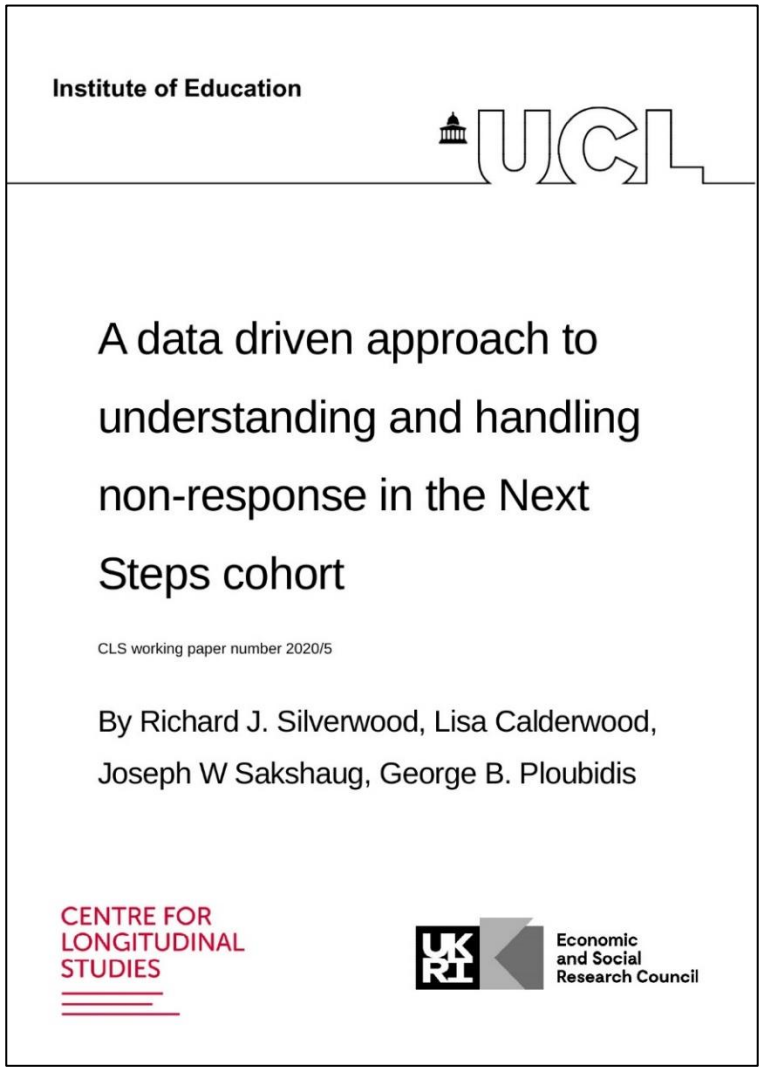
- We have identified variables which predict non-response at each sweep of BCS.
- These can be used as auxiliary variables in subsequent analyses to increase the plausibility of the MAR assumption.
- This approach can be extended for MNAR mechanisms in some cases, with appropriate external benchmark
- Simple test analyses have shown this approach to perform well.
- A straightforward approach, easily implemented in standard software.

CENTRE FOR LONGITUDINAL STUDIES

Other CLS studies

Next Steps

CENTRE FOR LONGITUDINAL STUDIES



Institute of Education

UCL

A data driven approach to understanding and handling non-response in the Next Steps cohort

CLS working paper number 2020/5

By Richard J. Silverwood, Lisa Calderwood, Joseph W Sakshaug, George B. Ploubidis

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

Millennium Cohort Study (MCS)

CENTRE FOR LONGITUDINAL STUDIES

- Ongoing work.

CENTRE FOR LONGITUDINAL STUDIES

COVID-19 Surveys

Non-response weights

CENTRE FOR LONGITUDINAL STUDIES

- 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.

Response

CENTRE FOR LONGITUDINAL STUDIES

	Wave 2			
Cohort	Issued sample	Response within issued sample	Target population	Response within target population
NSHD	2551	1569 (61.5%)	3758	1488 (39.6%)
NCDS	11,655	6282 (53.9%)	15,291	6228 (40.7%)
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%)**

Target population: individuals born in the specified birth period who are alive and still residing in the UK.

Overall response rate within issued sample (39.1%) comparable to similar COVID-19 web surveys.

Derivation of non-response weights

CENTRE FOR LONGITUDINAL STUDIES

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
Response model

CENTRE FOR LONGITUDINAL STUDIES

- 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.
- Full details in COVID-19 Surveys User Guide.

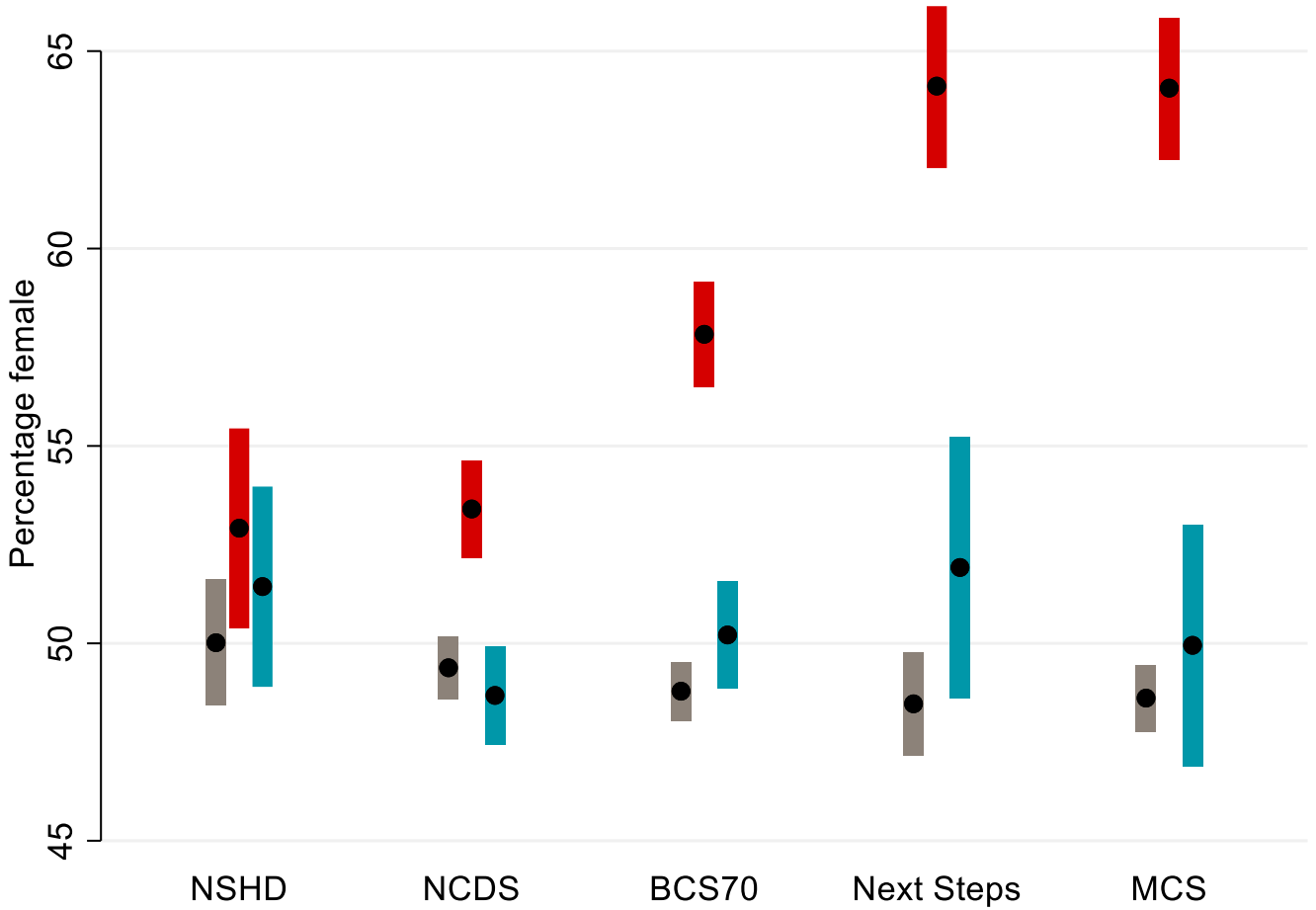
Derivation of non-response weights
Response model

CENTRE FOR LONGITUDINAL STUDIES

Sex	Internet access prior to web survey	Self-rated health
Ethnicity	Consent for biomarkers	Smoking status
Parental social class	Consent for linkages	Maternal mental health
Number of rooms at home/persons per room	Educational qualifications	Social capital/social support
Cognitive ability	Economic activity	Income
Early life mental health	Partnership status	Number of non-responses across all previous sweeps
Voting	Psychological distress	Response at COVID-19 Wave 1 survey*
Membership in organisations	BMI	

Effectiveness of non-response weights

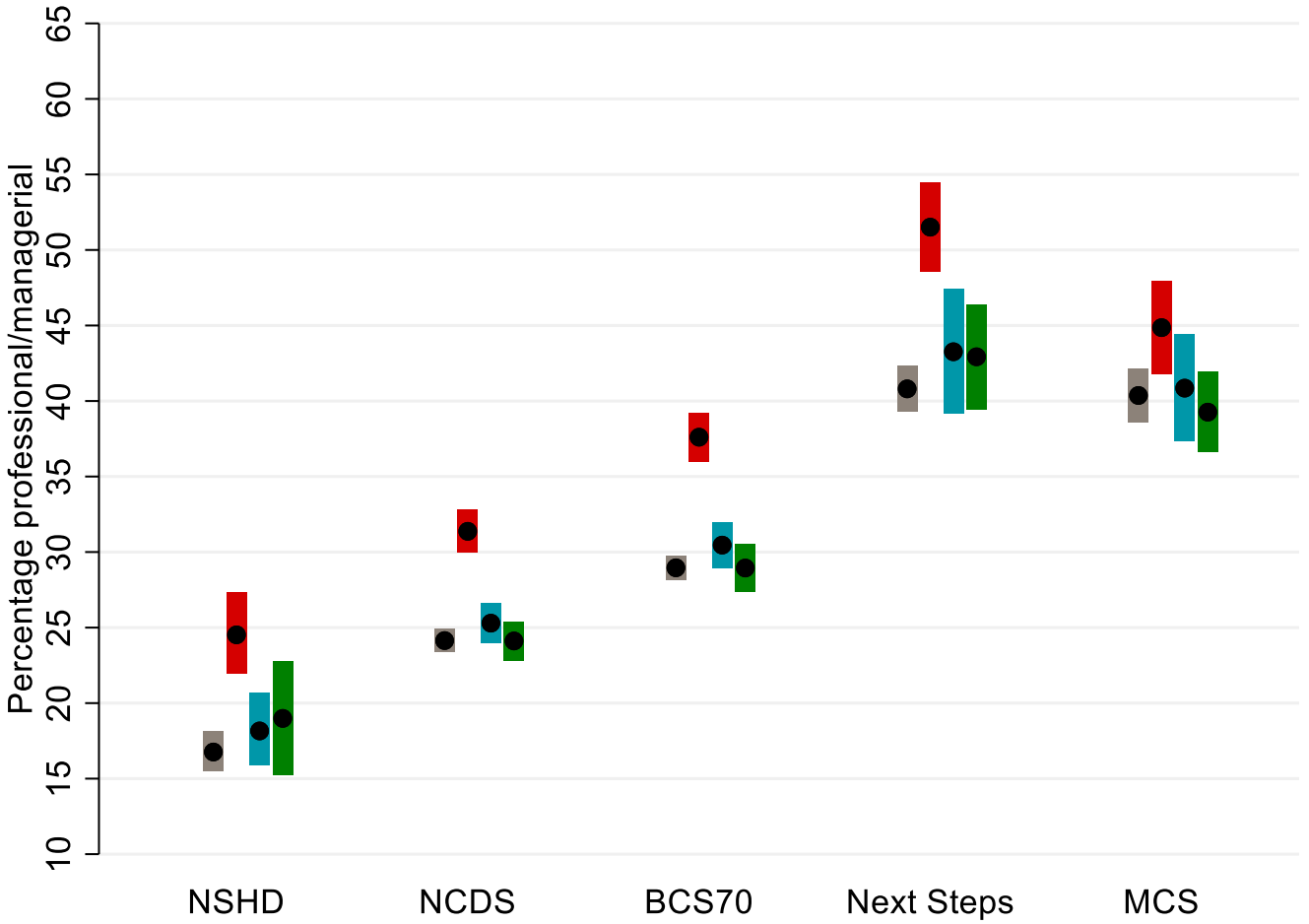
CENTRE FOR LONGITUDINAL STUDIES



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.

Comparison with MI (work in progress)

CENTRE FOR LONGITUDINAL STUDIES



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

CENTRE FOR LONGITUDINAL STUDIES

Leveraging data linkages

Extending the strategy

CENTRE FOR LONGITUDINAL STUDIES

- Growing interest in whether linked administrative data have the potential to aid analyses subject to missing data in cohort studies.
- Identify predictors of cohort non-response in linked administrative data.
- Explore whether added value in including identified variables as auxiliary variables with respect to restoring sample representativeness.
- Focusing on linked NCDS and hospital episode statistics (HES) data here. Many other linkages with CLS cohort data available.

Hospital Episode Statistics (HES)

CENTRE FOR LONGITUDINAL STUDIES

- A collection of databases containing details of interactions with NHS hospitals in England.
- Linkage between NCDS and HES datasets undertaken on the basis of consent at sweep 8 (age 50).
- Matching conducted using deterministic linkage based on combinations of the participant's name, sex, date of birth and postcode.
- Linked data available via secure access through the UK Data Service.

HES predictors of NCDS non-response

CENTRE FOR LONGITUDINAL STUDIES

- A total of 58 variables derived from HES data relating to:
  - Numbers of admissions and appointments
  - Missed appointments
  - Investigations undertaken
  - Diagnoses
  - Treatments received
- Employed a similar approach to identify most important predictors of NCDS non-response at wave 9 (age 55).
- 10 variables identified.

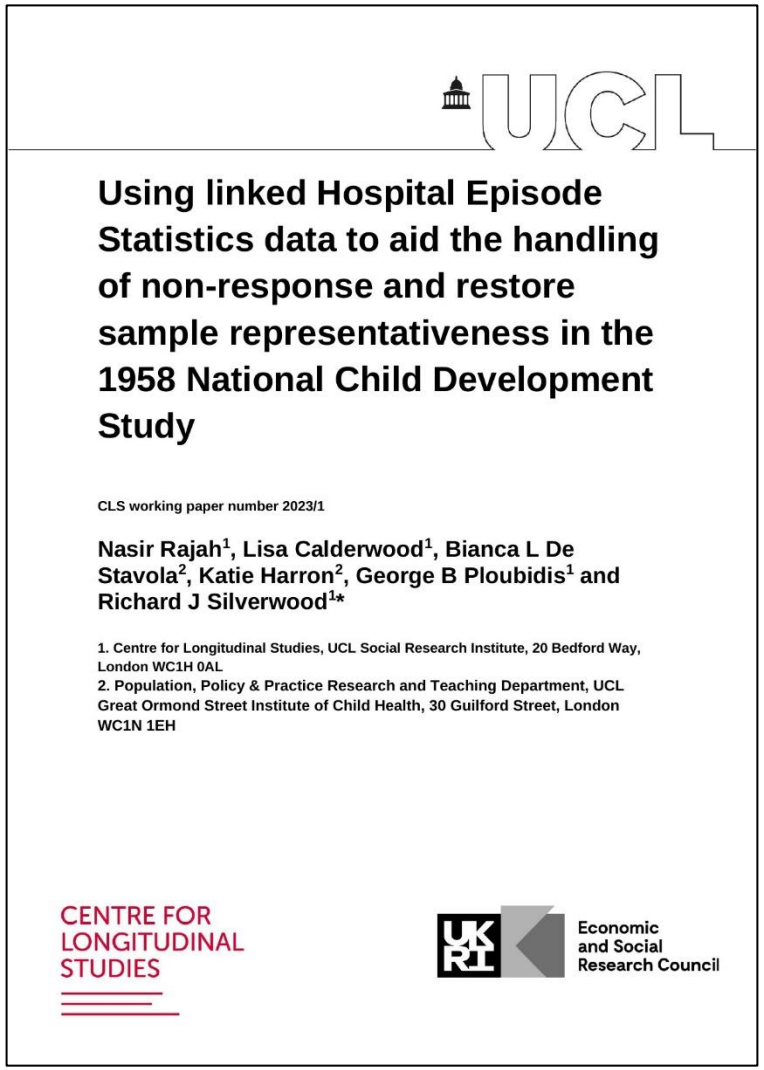
Restoring NCDS sample representativeness

CENTRE FOR LONGITUDINAL STUDIES

- Undertook similar test analysis to see if including the identified HES variables helped restore sample representative.
- Concluded that it did – a bit – but essentially no additional gain relative to using only previously identified survey predictors of non-response.

Leveraging data linkages

CENTRE FOR LONGITUDINAL STUDIES



UCL

Using linked Hospital Episode Statistics data to aid the handling of non-response and restore sample representativeness in the 1958 National Child Development Study

CLS working paper number 2023/1

Nasir Rajah[1], Lisa Calderwood[1], Bianca L De Stavola[2], Katie Harron[2], George B Ploubidis[1] and Richard J Silverwood[1]*

1. Centre for Longitudinal Studies, UCL Social Research Institute, 20 Bedford Way, London WC1H 0AL
2. Population, Policy & Practice Research and Teaching Department, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

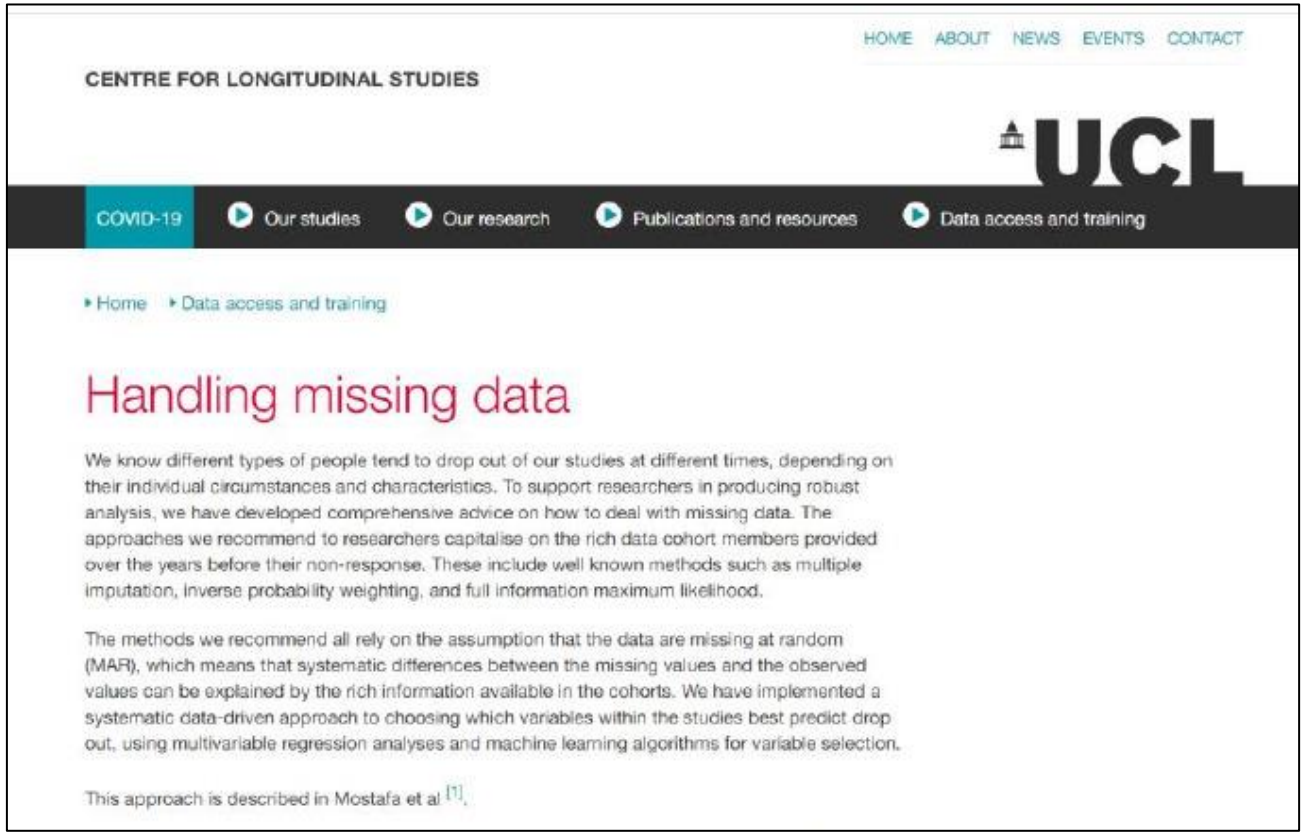
CENTRE FOR LONGITUDINAL STUDIES

Resources

Resources

CENTRE FOR LONGITUDINAL STUDIES

- Handling missing data webpage: https://cls.ucl.ac.uk/data-access-training/handling-missing-data/



Resources

CENTRE FOR LONGITUDINAL STUDIES

- Mostafa T, Narayanan M, Pongiglione B, Dodgeon B, Goodman A, Silverwood RJ, Ploubidis GB. Missing at random assumption made more plausible: evidence from the 1958 British birth cohort. J Clin Epidemiol. 2021;136:44-54.
- Silverwood RJ, Goodman A, Ploubidis GB. Letter to the editor: Don’t forget survey data: ‘healthy cohorts’ are ‘real-world’ relevant if missing data are handled appropriately. Longitudinal and Life Course Studies. 2022;13(2):335-41.
- Silverwood R, Narayanan M, Dodgeon B, Ploubidis G. Handling missing data in the National Child Development Study: User Guide (Version 2). London: UCL Centre for Longitudinal Studies; 2021.

Resources

CENTRE FOR LONGITUDINAL STUDIES

- 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.
- Brown M, Goodman A, Peters A, Ploubidis GB, Sanchez A, Silverwood R, et al. COVID-19 Survey in Five National Longitudinal Studies: Waves 1, 2 and 3 User Guide (Version 3). London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing; 2021.
- Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB, Silverwood RJ. Using linked Hospital Episode Statistics data to better handle non-response and restore sample representativeness in the National Child Development Study. CLS Working Paper Series 2023/1. London: UCL Centre for Longitudinal Studies; 2023.

Institute of Education

UCL

Thank you.

CENTRE FOR
LONGITUDINAL
STUDIES