Examples of the types of analyses that can be undertaken using CLS cohort data

Richard Silverwood
Associate Professor of Statistics &
CLS Chief Statistician

29 November 2022

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

Outline

CENTRE FOR LONGITUDINAL STUDIES

1. Simple analyses
2. Confounder control
3. Repeated measures
4. Cross-cohort analysis

Simple analyses

Simple analyses

CENTRE FOR
LONGITUDINAL
STUDIES

- The cohorts provide rich data collected on cohort members over many years/decades, so complex analyses possible.
- But let's start with some simple examples…

Simple analyses: Examples

CENTRE FOR
LONGITUDINAL
STUDIES

UCL

Prevalence of COVID-19, symptoms and testing in the U.K.

Initial findings from the COVID-19 Survey in Five National Longitudinal Studies

By Dylan M. Williams, Gabriella Conti, Nishi Chaturvedi, Alun Hughes, George B. Ploubidis and Richard J. Silverwood

Simple analyses: Examples

CENTRE FOR LONGITUDINAL STUDIES

What is the prevalence of test-confirmed COVID-19 in each cohort?

COVIDTEST Have you been tested for Coronavirus?

- Yes
- No

COVIDRESULT What was the result of your coronavirus test? If you had more than one test please report the findings of the latest test.

- Positive - it showed I had coronavirus
- Negative - it showed I did not have coronavirus
- Inconclusive
- Waiting for results

Table 1 – Percentages of respondents that had tested positive for SARS-CoV-2 by cohort

Cohort	n	N	Percent	95% CI
MCS	8	2609	0.3	0.1, 0.8
NS	10	1876	0.6	0.3, 1.1
BCS70	28	4132	0.7	0.5, 1.0
NCDS	15	5119	0.3	0.2, 0.5
NSHD	1	1170	0.1	0.0, 0.5

n – number reporting a positive test; N – total sample size

CENTRE FOR
LONGITUDINAL
STUDIES

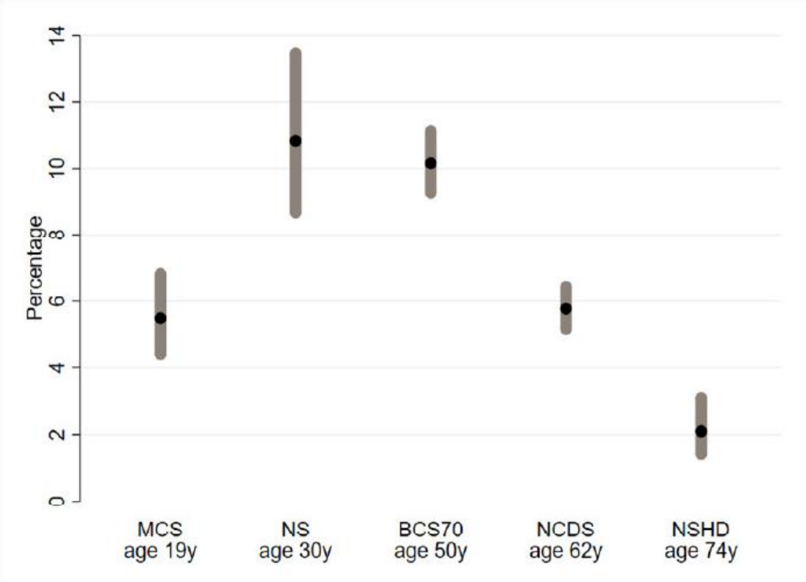
Simple analyses: Examples

What is the prevalence of self-reported COVID-19 in each cohort?

COVID19 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

Figure 1 – Prevalences of confirmed or self-reported COVID-19 by cohort



Ranges show point estimates with 95% CI

Simple analyses: Examples

CENTRE FOR LONGITUDINAL STUDIES

Does the prevalence of self-reported COVID-19 differ between males and females in each cohort?

Cohort	Male/female	N (total)	N (C-19) (%)	Risk ratio	95% CI
MCS	Male	770	38 (4.9)	1.00	(ref)
	Female	1,839	120 (6.5)	1.45	0.92, 2.27
Next Steps	Male	643	82 (12.8)	1.00	(ref)
	Female	1,233	115 (9.3)	0.68	0.44, 1.06
BCS70	Male	1,711	160 (9.4)	1.00	(ref)
	Female	2,420	219 (9.0)	1.00	0.71, 1.41
NCDS	Male	2,432	137 (5.6)	1.00	(ref)
	Female	2,686	159 (5.9)	1.17	0.83, 1.66

Confounder control

Confounder control

CENTRE FOR LONGITUDINAL STUDIES

- If we want an estimated association between an independent variable and a dependent variable to have any causal interpretation, we need to consider confounder control.
- **Confounder**: A variable that causes non-causal (spurious) association between an independent variable and a dependent variable.

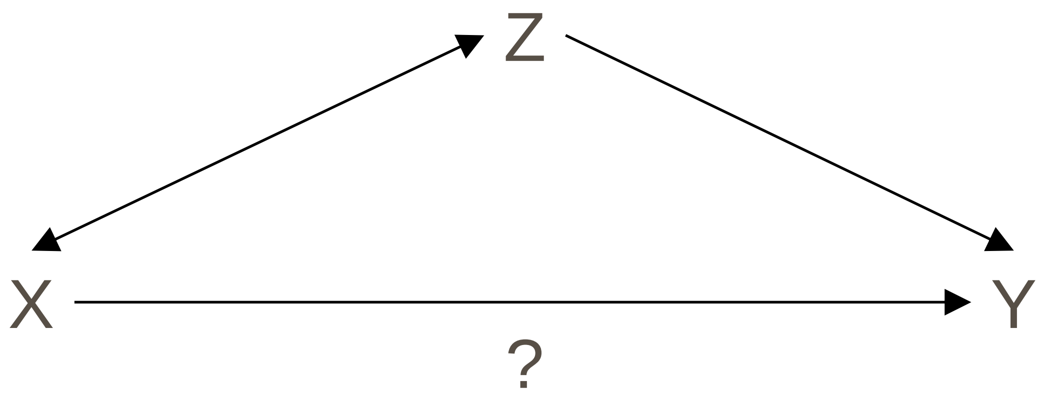
Confounder control

- If we want an estimated association between an independent variable and a dependent variable to have any causal interpretation, we need to consider confounder control.
- **Confounder**: A variable that causes non-causal (spurious) association between an independent variable and a dependent variable.

$$X \xrightarrow[?]{} Y$$

Confounder control

- If we want an estimated association between an independent variable and a dependent variable to have any causal interpretation, we need to consider confounder control.
- **Confounder**: A variable that causes non-causal (spurious) association between an independent variable and a dependent variable.



Confounder control

CENTRE FOR LONGITUDINAL STUDIES

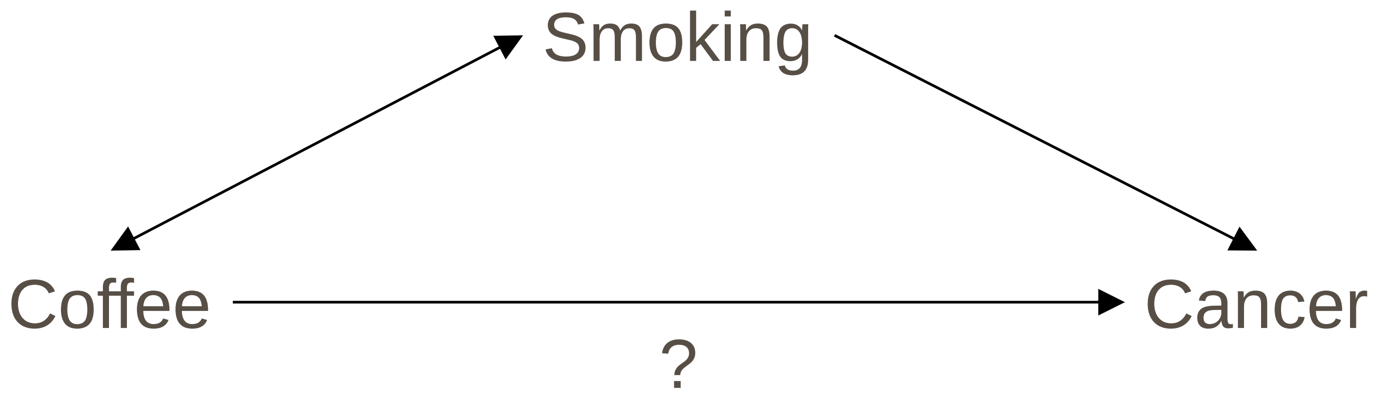
- If we want an estimated association between an independent variable and a dependent variable to have any causal interpretation, we need to consider confounder control.
- **Confounder**: A variable that causes non-causal (spurious) association between an independent variable and a dependent variable.

$$\text{Coffee} \xrightarrow[?]{} \text{Cancer}$$

Confounder control

CENTRE FOR LONGITUDINAL STUDIES

- If we want an estimated association between an independent variable and a dependent variable to have any causal interpretation, we need to consider confounder control.
- **Confounder**: A variable that causes non-causal (spurious) association between an independent variable and a dependent variable.



Confounder control

- If we want an estimated association between an independent variable and a dependent variable to have any causal interpretation, we need to consider confounder control.
- **Confounder**: A variable that causes non-causal (spurious) association between an independent variable and a dependent variable.
- Thankfully, the rich data collected on cohort members over many years/decades provide great opportunity for confounder control.

Confounder control: Example

CENTRE FOR
LONGITUDINAL
STUDIES

Research

JAMA Psychiatry | Original Investigation

Association of Early-Life Mental Health With Biomarkers in Midlife and Premature Mortality
Evidence From the 1958 British Birth Cohort

George B. Ploubidis, PhD; G. David Batty, PhD, DSc; Praveetha Patalay, PhD;
David Bann, PhD; Alissa Goodman, MSc

Supplemental content

IMPORTANCE Early-life mental health is known to be associated with socioeconomic adversity and psychological distress in adulthood, but less is known about potential associations with biomarkers and mortality.

OBJECTIVE To investigate the association between early-life mental health trajectories with biomarkers in midlife and premature mortality.

DESIGN, SETTING, AND PARTICIPANTS This study used data from the British National Child Development Study, a population-based birth cohort. The initial sample of 17 415 individuals consisted of all infants born in Great Britain in a single week in 1958. Analysis began Feburary 2017 and ended May 2020.

Confounder control: Example

CENTRE FOR LONGITUDINAL STUDIES

Early-life mental health (age 7-16) → Biomarkers in midlife (age 44-45)

?

Confounder control: Example

CENTRE FOR LONGITUDINAL STUDIES

Early-life mental health (age 7-16)
Rutter Child Scale A at ages 7 and 11 (mothers) and at age 16 (teachers):
- Conduct problems
- Affective symptoms

Early-life mental health (age 7-16) → ? → Biomarkers in midlife (age 44-45)

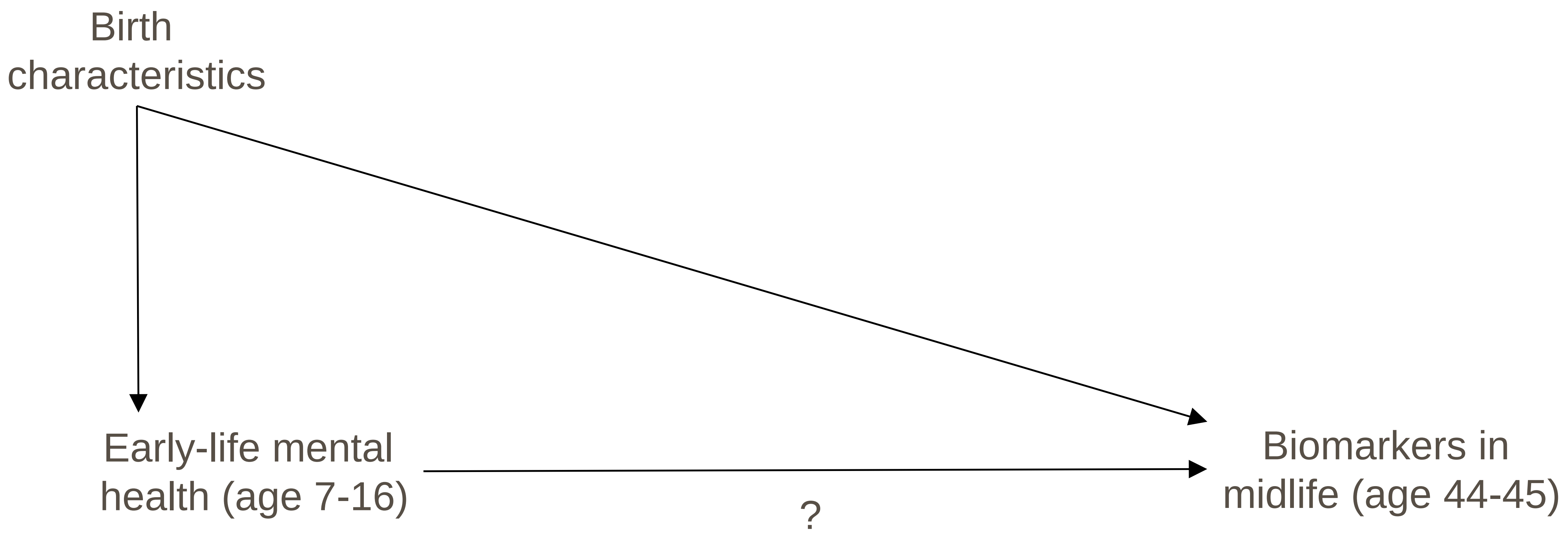
Confounder control: Example

Biomarkers in midlife (age 44-45)
- Fibrinogen
- C-reactive protein
- Glycated haemoglobin
- High-density lipoprotein
- Low-density lipoprotein
- High blood pressure

Early-life mental health (age 7-16) → Biomarkers in midlife (age 44-45)

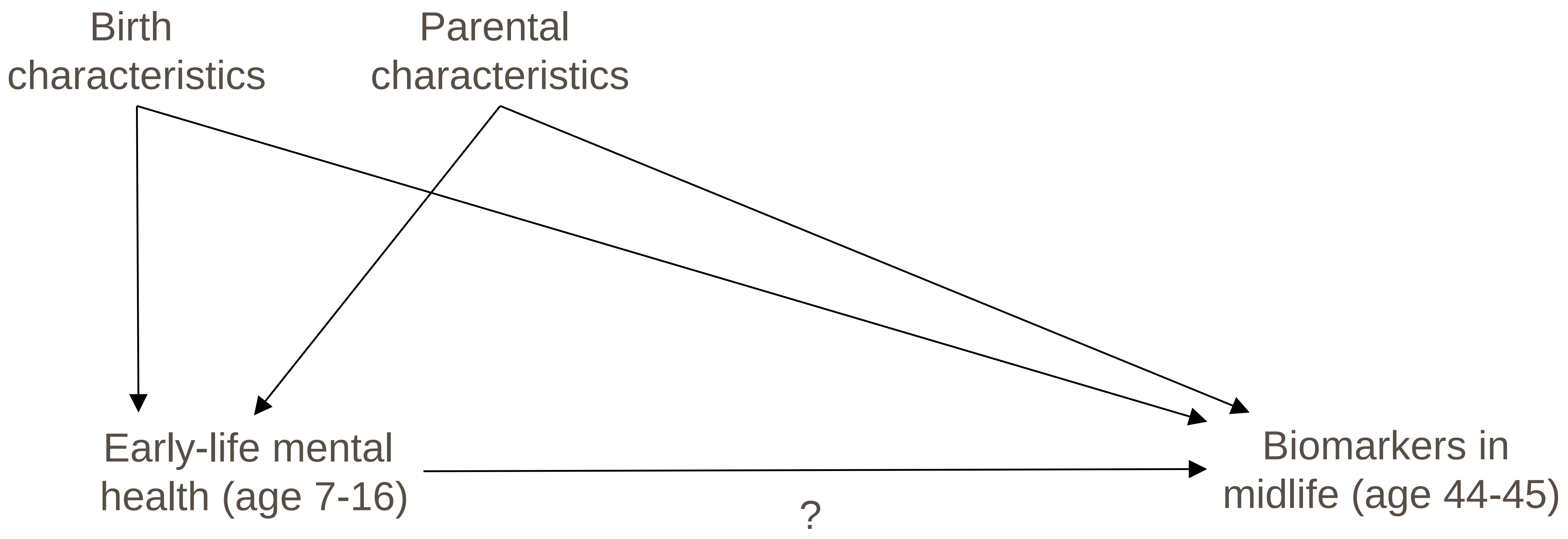
?

Confounder control: Example



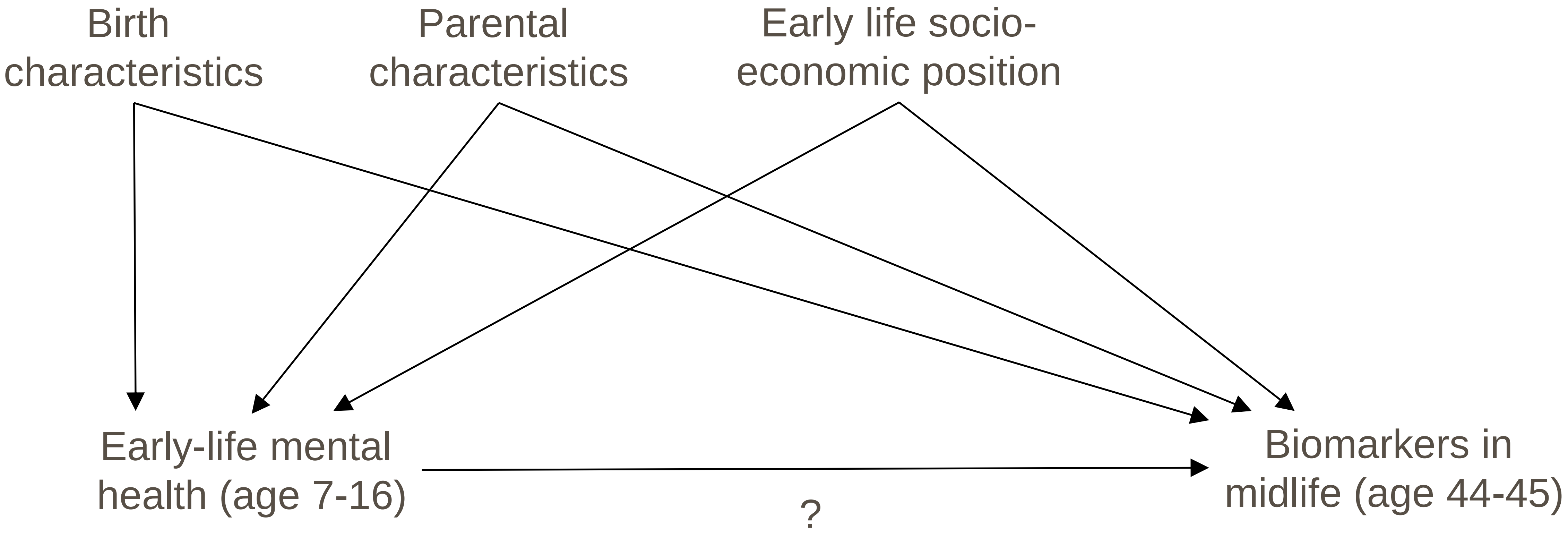
Confounder control: Example

CENTRE FOR LONGITUDINAL STUDIES



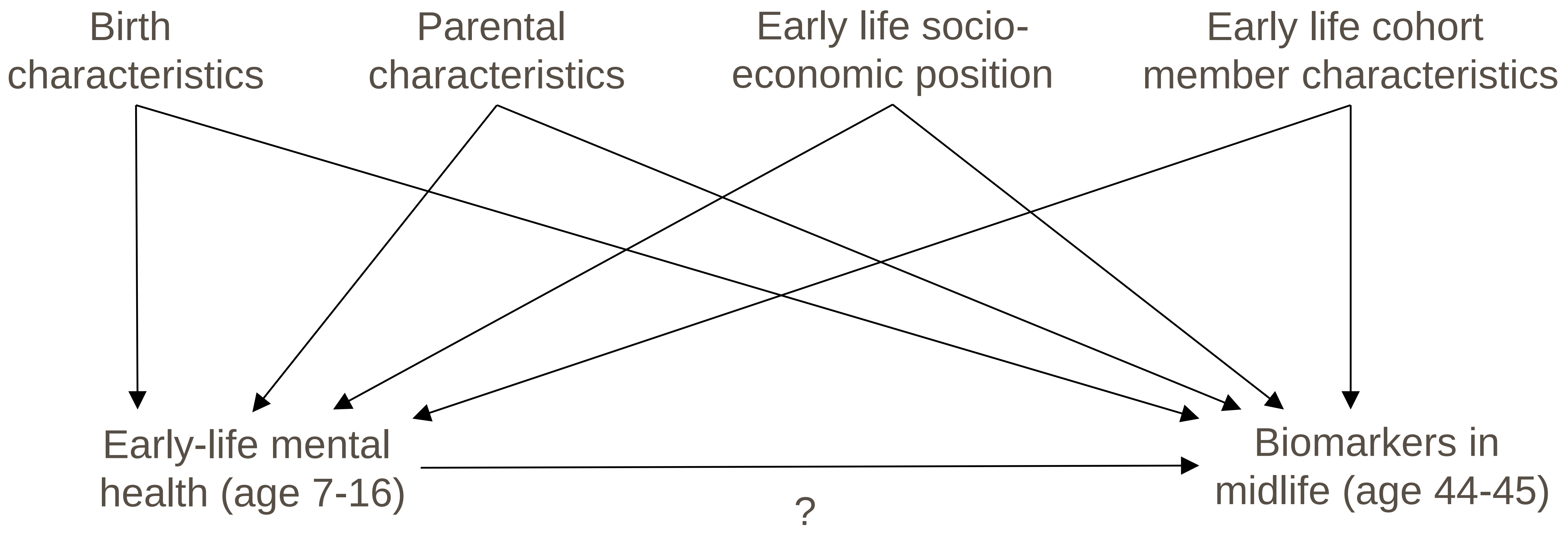
Confounder control: Example

CENTRE FOR LONGITUDINAL STUDIES



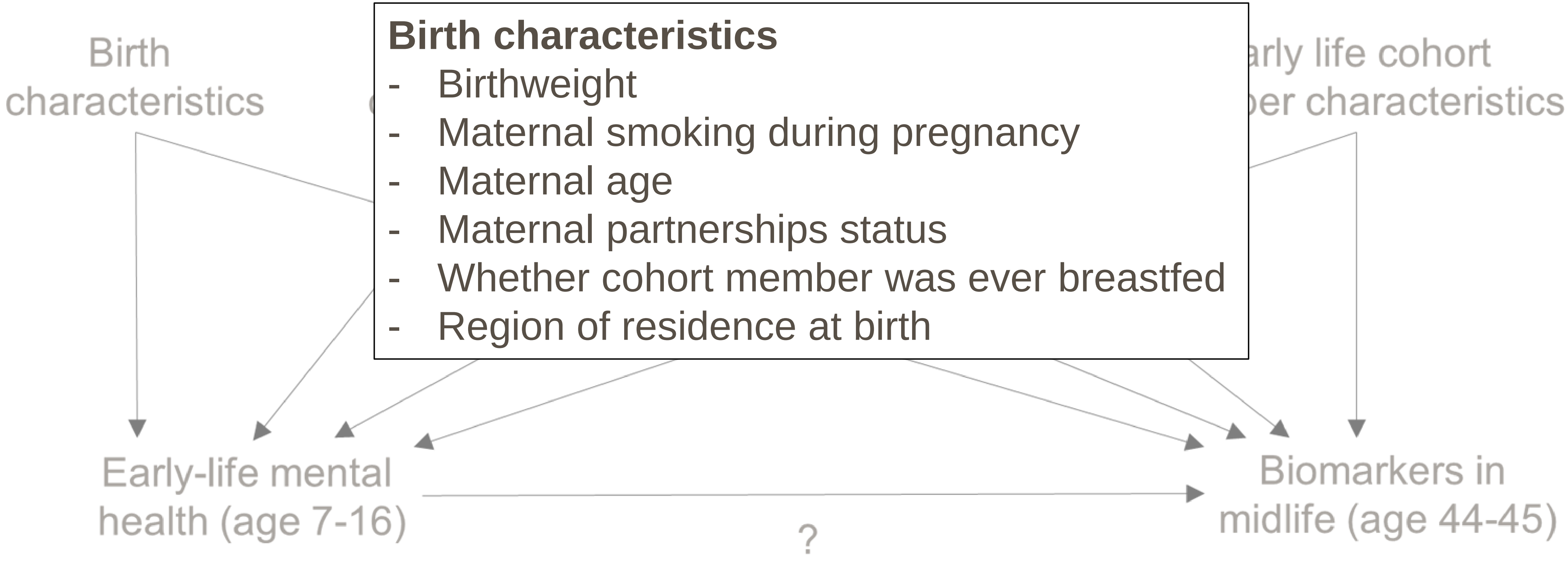
Confounder control: Example

CENTRE FOR LONGITUDINAL STUDIES



CENTRE FOR
LONGITUDINAL
STUDIES

Confounder control: Example

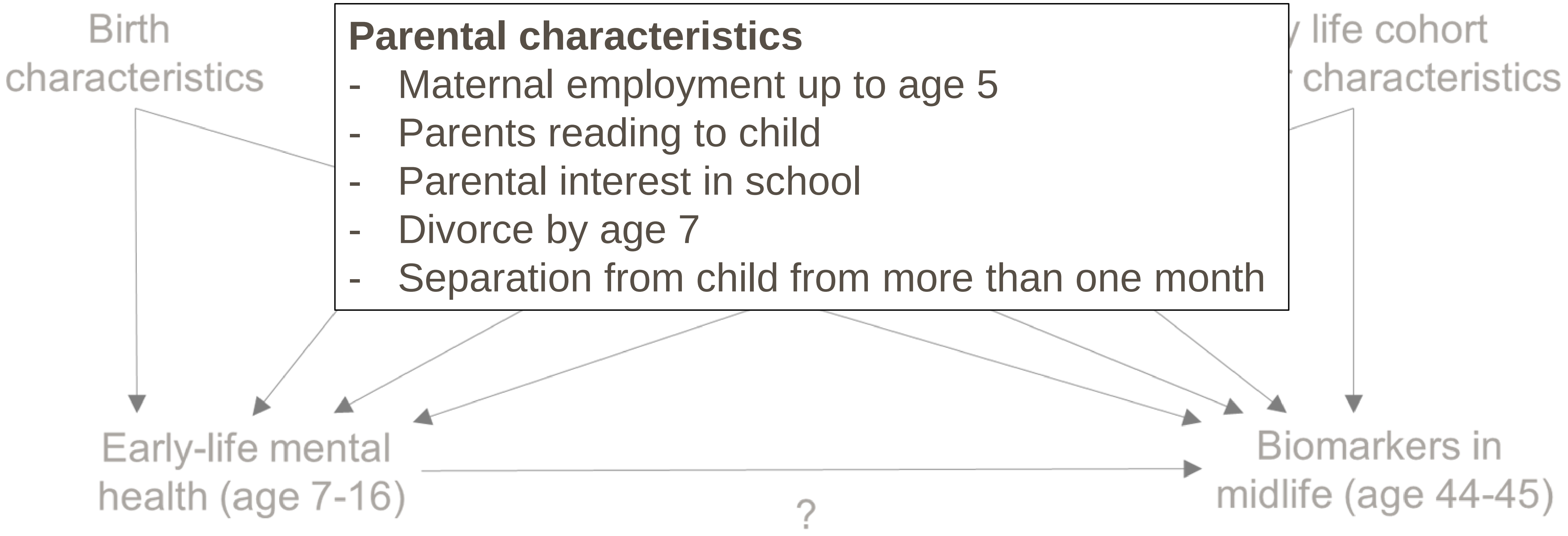


Birth characteristics

- Birthweight
- Maternal smoking during pregnancy
- Maternal age
- Maternal partnerships status
- Whether cohort member was ever breastfed
- Region of residence at birth

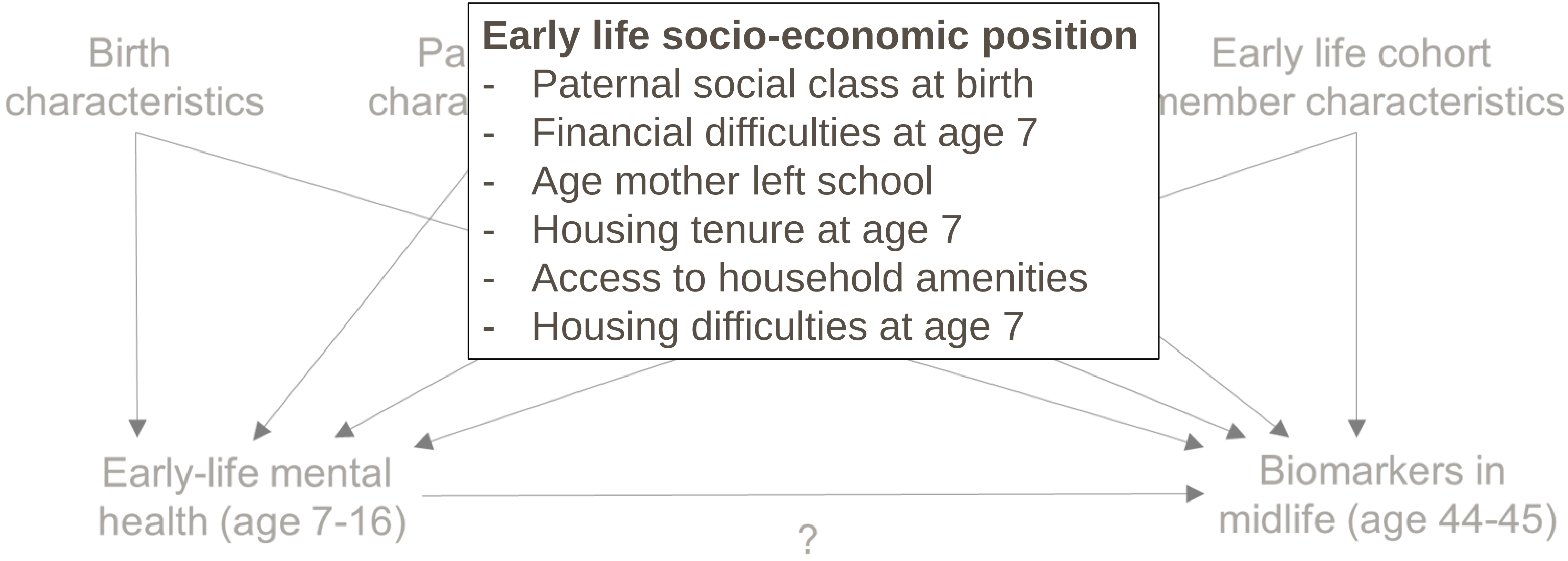
Confounder control: Example

CENTRE FOR
LONGITUDINAL
STUDIES



Confounder control: Example

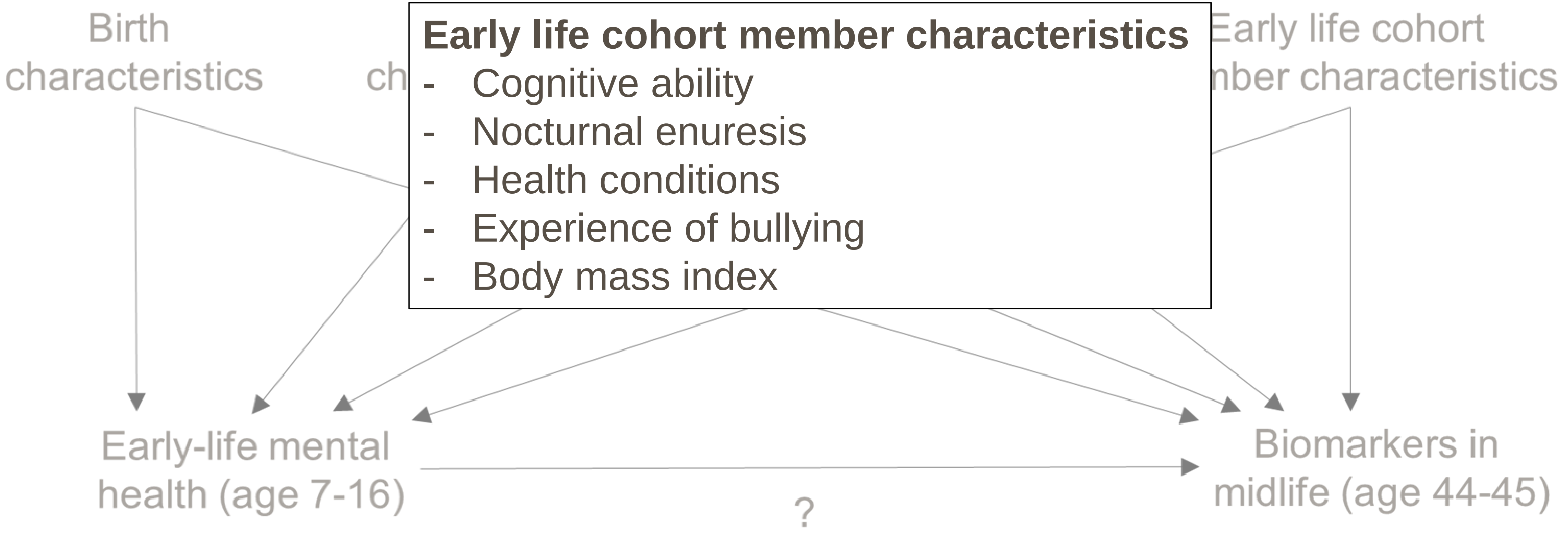
CENTRE FOR LONGITUDINAL STUDIES



Early life socio-economic position

- Paternal social class at birth
- Financial difficulties at age 7
- Age mother left school
- Housing tenure at age 7
- Access to household amenities
- Housing difficulties at age 7

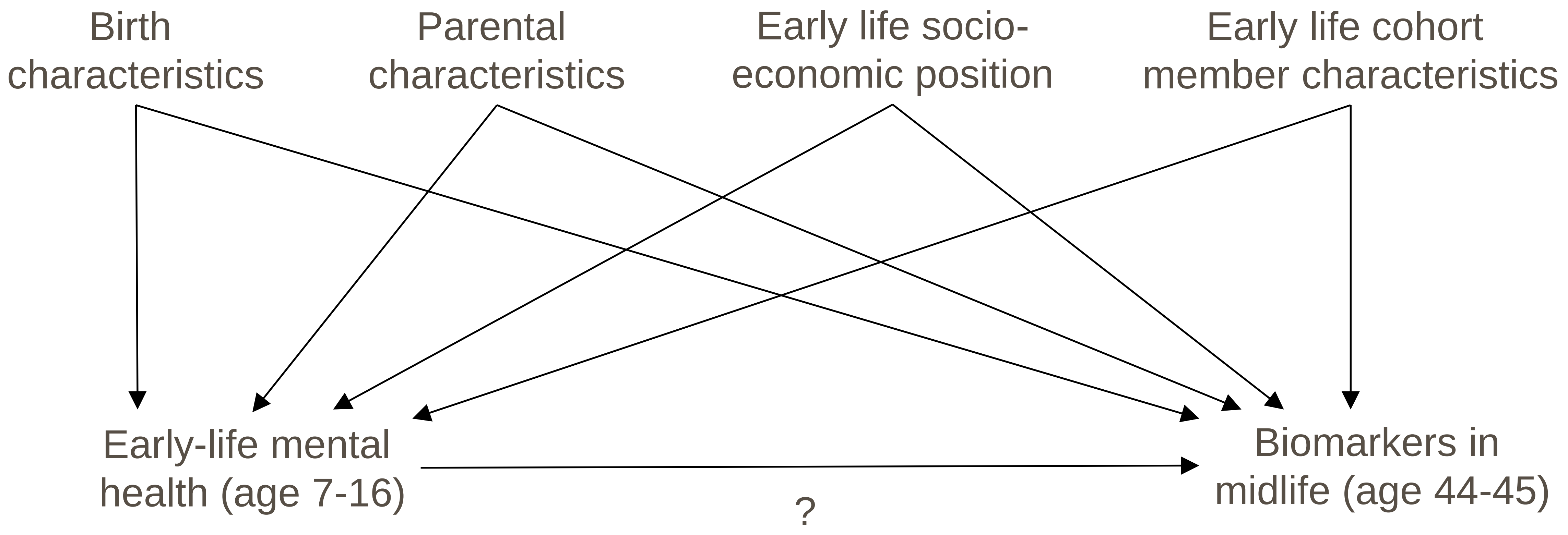
Confounder control: Example



Early life cohort member characteristics

- Cognitive ability
- Nocturnal enuresis
- Health conditions
- Experience of bullying
- Body mass index

Confounder control: Example



Confounder control: Key message

- The rich data collected on cohort members over many years/decades provide great opportunity for confounder control.

Repeated measures

Repeated measures

CENTRE FOR LONGITUDINAL STUDIES

- Long-running cohorts measuring consistent topics over time provide repeated measures of the same measurement/construct.
- Examples:
  - Physical measurements
  - General physical health, mental health, specific diseases/conditions, health behaviours
  - Relationships, marital status, household composition
  - Employment status, occupation, earnings and income
- Allows you to characterise *changes* or *trajectories* over time.

Repeated measures: Example

CENTRE FOR LONGITUDINAL STUDIES

Research

JAMA Dermatology | Original Investigation

Patterns of Atopic Eczema Disease Activity From Birth Through Midlife in 2 British Birth Cohorts

Katrina Abuabara, MD, MA, MSCE; Morgan Ye, MPH; David J. Margolis, MD, PhD; Charles E. McCulloch, PhD;
Amy R. Mulick, MSc; Richard J. Silverwood, PhD; Alice Sullivan, PhD;
Hywel C. Williams, DSc; Sinéad M. Langan, PhD

Supplemental content

IMPORTANCE Atopic eczema is characterized by a heterogenous waxing and waning course, with variable age of onset and persistence of symptoms. Distinct patterns of disease activity such as early-onset/resolving and persistent disease have been identified throughout childhood; little is known about patterns into adulthood.

OBJECTIVE This study aimed to identify subtypes of atopic eczema based on patterns of disease activity through mid-adulthood, to examine whether early life risk factors and participant characteristics are associated with these subtypes, and to determine whether subtypes are associated with other atopic diseases and general health in mid-adulthood.

DESIGN, SETTING, AND PARTICIPANTS This study evaluated members of 2 population-based birth cohorts, the 1958 National Childhood Development Study (NCDS) and the 1970 British Cohort Study (BCS70). Participant data were collected over the period between 1958 and 2016. Data were analyzed over the period between 2018 and 2020.

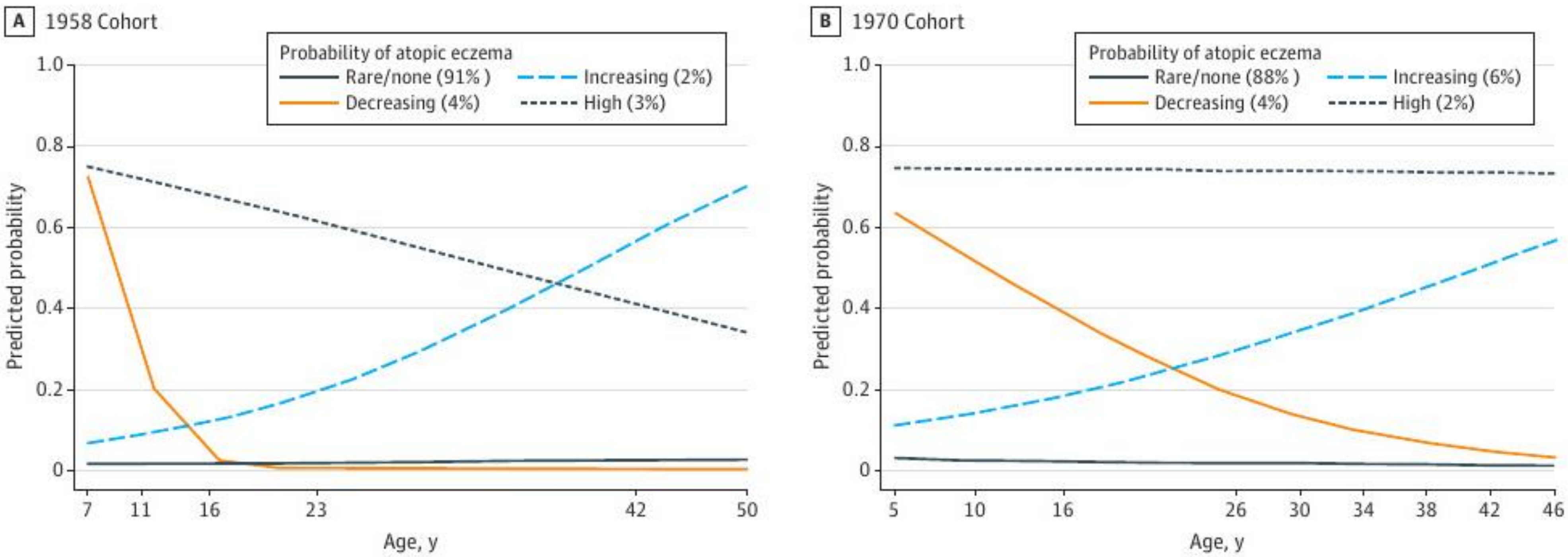
Repeated measures: Example

- Aimed to identify subtypes of eczema based on patterns of disease activity in NCDS and BCS70.
- Parent-reported or self-reported eczema period prevalence available from standardised questions at ages 7, 11, 16, 23, 42 and 50 in NCDS and ages 5, 10, 16, 26, 30, 34, 38, 42 and 46 in BCS70.
- Then examined whether:
  - early life risk factors associated with eczema subtypes
  - eczema subtypes associated with other atopic diseases and general health in mid-adulthood

Repeated measures: Example

CENTRE FOR
LONGITUDINAL
STUDIES

Figure. Estimated Probabilities of Atopic Eczema Symptoms at Each Age for Each Subtype in 4 Class Models From 2 British Birth Cohorts



Predicted probabilities at each age generated from generalized linear and latent mixed models.

Repeated measures: Key message

CENTRE FOR LONGITUDINAL STUDIES

- British cohort studies provide repeated observations of the same measurement/construct.
- Allows you to characterise changes or trajectories over time.

Cross-cohort analysis

Cross-cohort analysis

CENTRE FOR LONGITUDINAL STUDIES

- Conducting analyses across multiple cohorts allows us to extend our hypotheses: how do things change over time or between cohorts?
- Ideally want to analyse *identical* measures across cohorts.
- In absence of this, need consider how measures can best be *harmonised*.
- COVID-19 surveys offer great opportunity for cross-cohort analysis as most questions identical.

CENTRE FOR LONGITUDINAL STUDIES

Cross-cohort analysis: Example

CrossMark

Socioeconomic inequalities in childhood and adolescent body-mass index, weight, and height from 1953 to 2015: an analysis of four longitudinal, observational, British birth cohort studies

David Bann, William Johnson, Leah Li, Diana Kuh, Rebecca Hardy

oa OPEN ACCESS

Summary

Background Socioeconomic inequalities in childhood body-mass index (BMI) have been documented in high-income countries; however, uncertainty exists with regard to how they have changed over time, how inequalities in the composite parts (ie, weight and height) of BMI have changed, and whether inequalities differ in magnitude across the outcome distribution. Therefore, we aimed to investigate how socioeconomic inequalities in childhood and adolescent weight, height, and BMI have changed over time in Britain.

Methods We used data from four British longitudinal, observational, birth cohort studies: the 1946 Medical Research Council National Survey of Health and Development (1946 NSHD), 1958 National Child Development Study (1958 NCDS), 1970 British Cohort Study (1970 BCS), and 2001 Millennium Cohort Study (2001 MCS). BMI (kg/m^2) was derived in each study from measured weight and height. Childhood socioeconomic position was indicated by the

Lancet Public Health 2018; 3: e194–203

Published Online March 20, 2018 http://dx.doi.org/10.1016/S2468-2667(18)30045-8

See **Editorial** page e153

See **Comment** page e160

Centre for Longitudinal Studies, University College London (UCL) Institute of

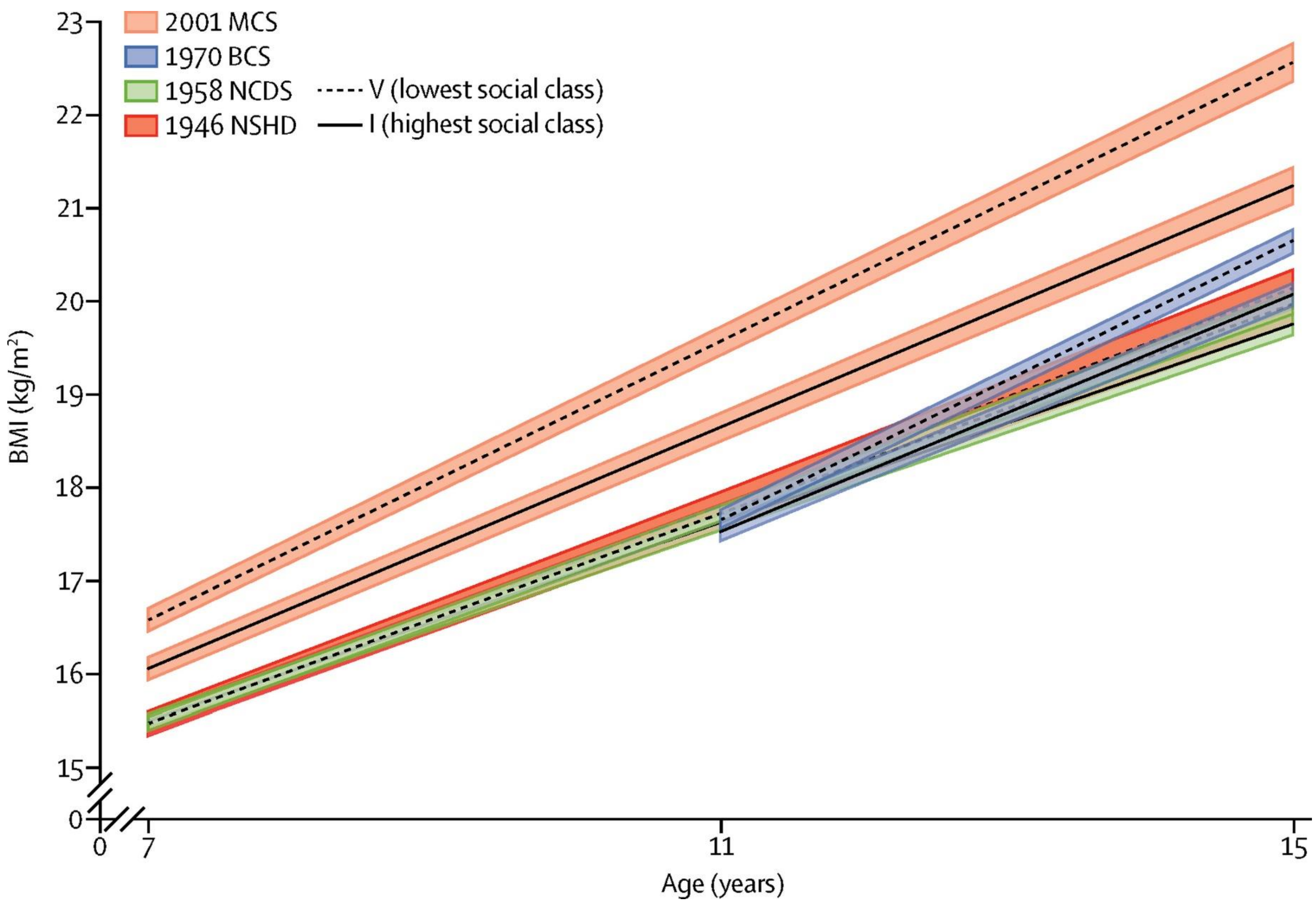
Cross-cohort analysis: Example

CENTRE FOR LONGITUDINAL STUDIES

- Investigated how socioeconomic inequalities in childhood and adolescent weight, height, and BMI have changed over time.
- Used data from NSHD (BMI at ages 7, 11 and 15), NCDS (7, 11 and 16), BCS70 (10 and 16) and MCS (7, 11 and 14).
- Childhood socioeconomic position indicated by father's occupational social class reported at age 10-11.
- Examined associations between childhood socioeconomic position and BMI to assess socioeconomic inequalities.
- Examined whether inequalities widened or narrowed from childhood to adolescence.

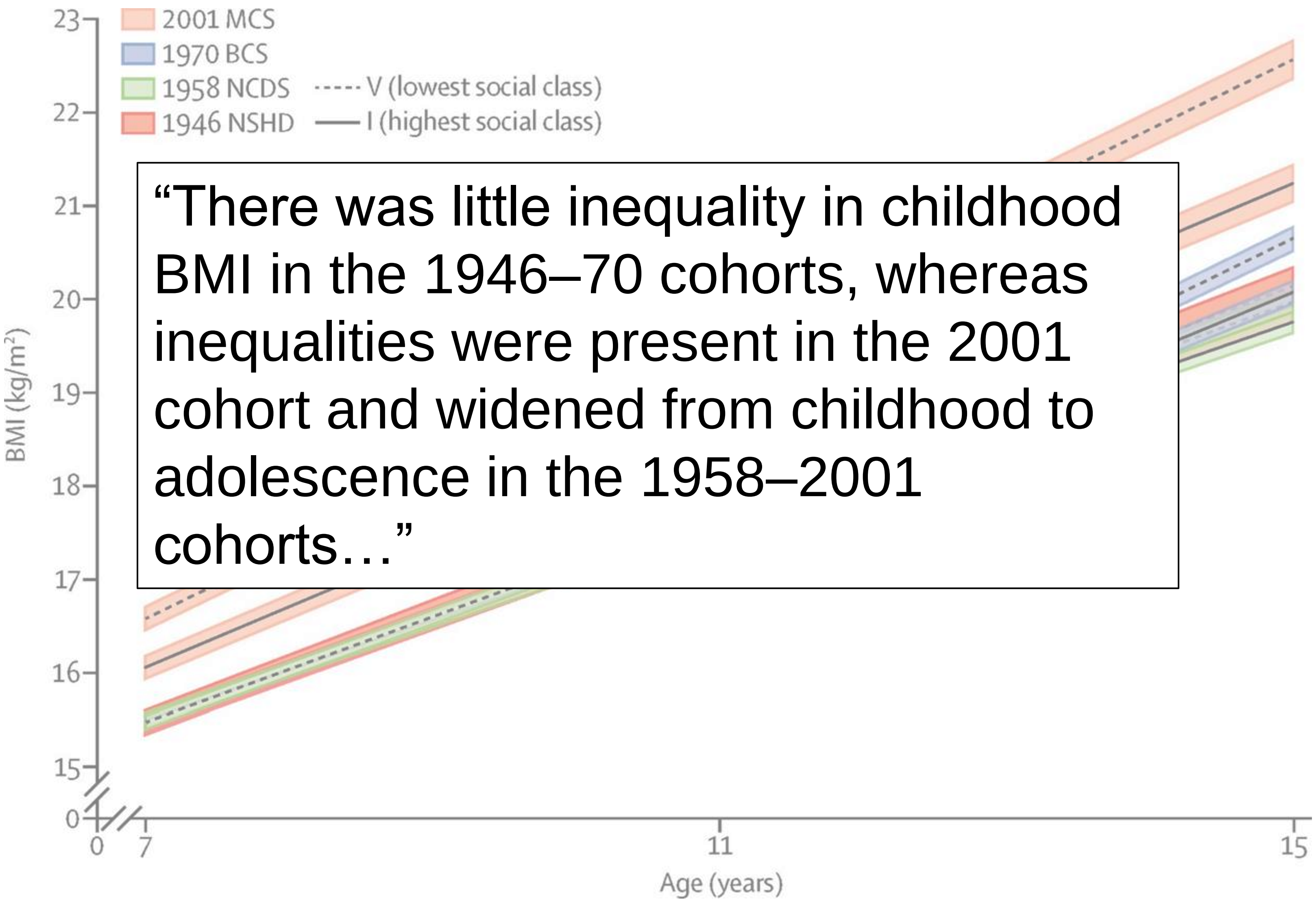
Cross-cohort analysis: Example

CENTRE FOR LONGITUDINAL STUDIES



Cross-cohort analysis: Example

CENTRE FOR LONGITUDINAL STUDIES



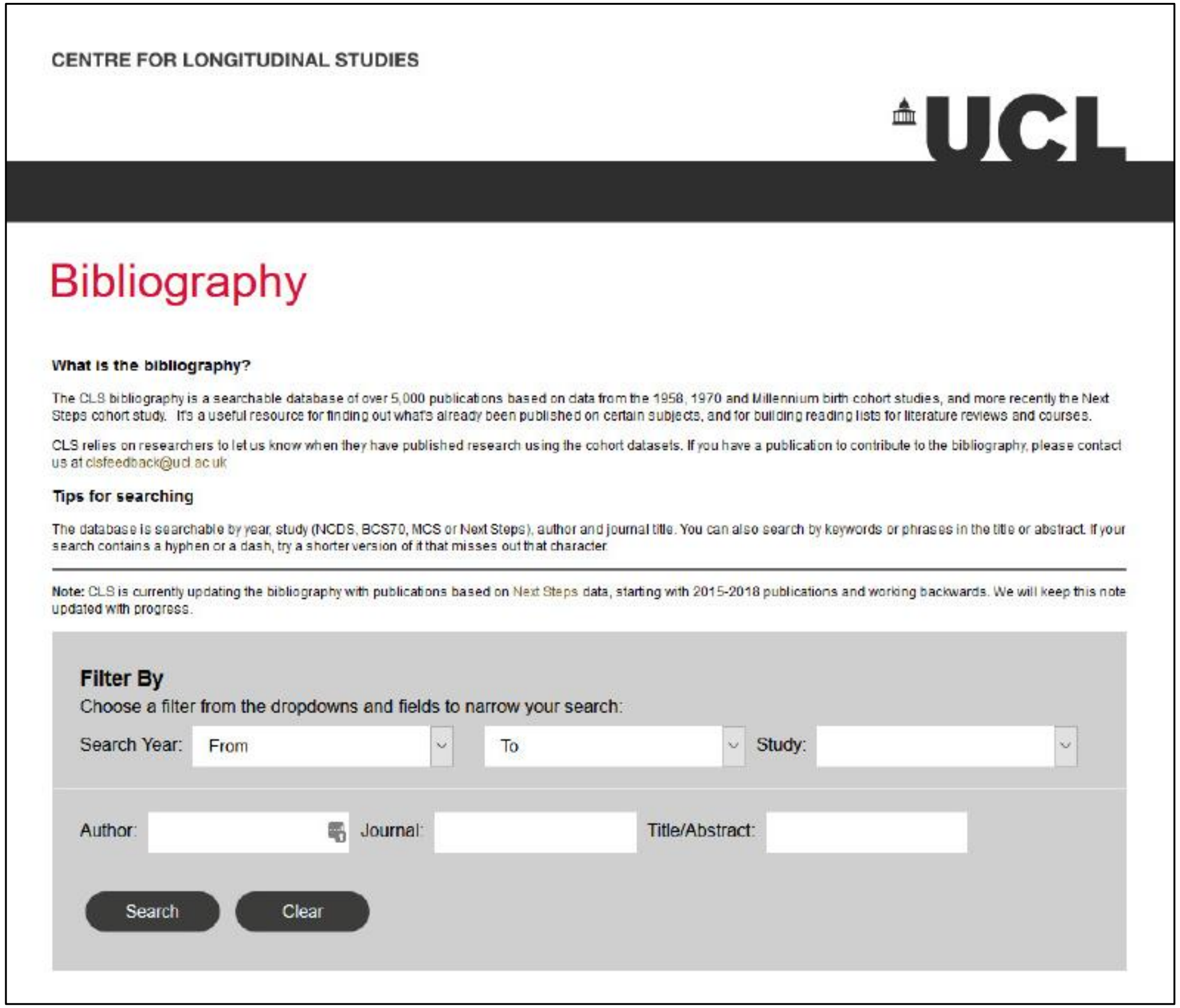
"There was little inequality in childhood BMI in the 1946–70 cohorts, whereas inequalities were present in the 2001 cohort and widened from childhood to adolescence in the 1958–2001 cohorts…"

Cross-cohort analysis: Key message

- Conducting analyses across multiple cohorts allows us to extend our hypotheses: how do things change over time or between cohorts?

CLS bibliography

CENTRE FOR LONGITUDINAL STUDIES


CENTRE FOR LONGITUDINAL STUDIES

UCL

Bibliography

What is the bibliography?

The CLS bibliography is a searchable database of over 5,000 publications based on data from the 1958, 1970 and Millennium birth cohort studies, and more recently the Next Steps cohort study. It's a useful resource for finding out what's already been published on certain subjects, and for building reading lists for literature reviews and courses.

CLS relies on researchers to let us know when they have published research using the cohort datasets. If you have a publication to contribute to the bibliography, please contact us at clsfeedback@ucl.ac.uk

Tips for searching

The database is searchable by year, study (NCDS, BCS70, MCS or Next Steps), author and journal title. You can also search by keywords or phrases in the title or abstract. If your search contains a hyphen or a dash, try a shorter version of it that misses out that character.

Note: CLS is currently updating the bibliography with publications based on Next Steps data, starting with 2015-2018 publications and working backwards. We will keep this note updated with progress.

Filter By

Choose a filter from the dropdowns and fields to narrow your search:

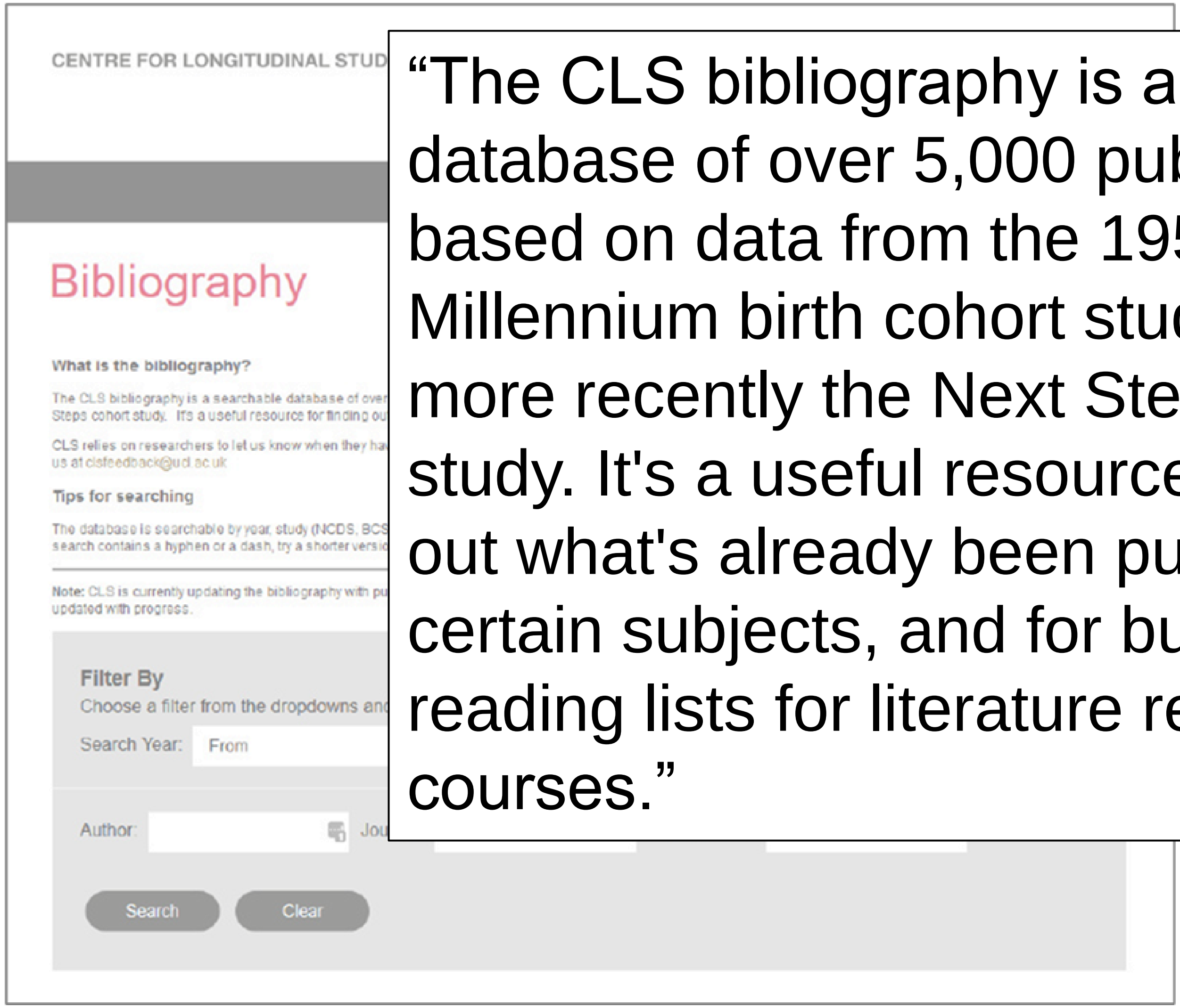
Search Year: From To Study:

Author: Journal: Title/Abstract:

Search Clear

https://www.bibliography.cls.ucl.ac.uk

CLS bibliography



"The CLS bibliography is a searchable database of over 5,000 publications based on data from the 1958, 1970 and Millennium birth cohort studies, and more recently the Next Steps cohort study. It's a useful resource for finding out what's already been published on certain subjects, and for building reading lists for literature reviews and courses."

w.bibliography.
cls.ucl.ac.uk

Thank you.

CENTRE FOR
LONGITUDINAL
STUDIES

Economic
and Social
Research Council