

Restructuring the Millennium Cohort Study data sets for longitudinal analysis in R: An Open Science project



Lydia Gabriela Speyer & Hildigunnur Anna Hall
Department of Psychology, University of Edinburgh

Millennium Cohort Study (MCS)

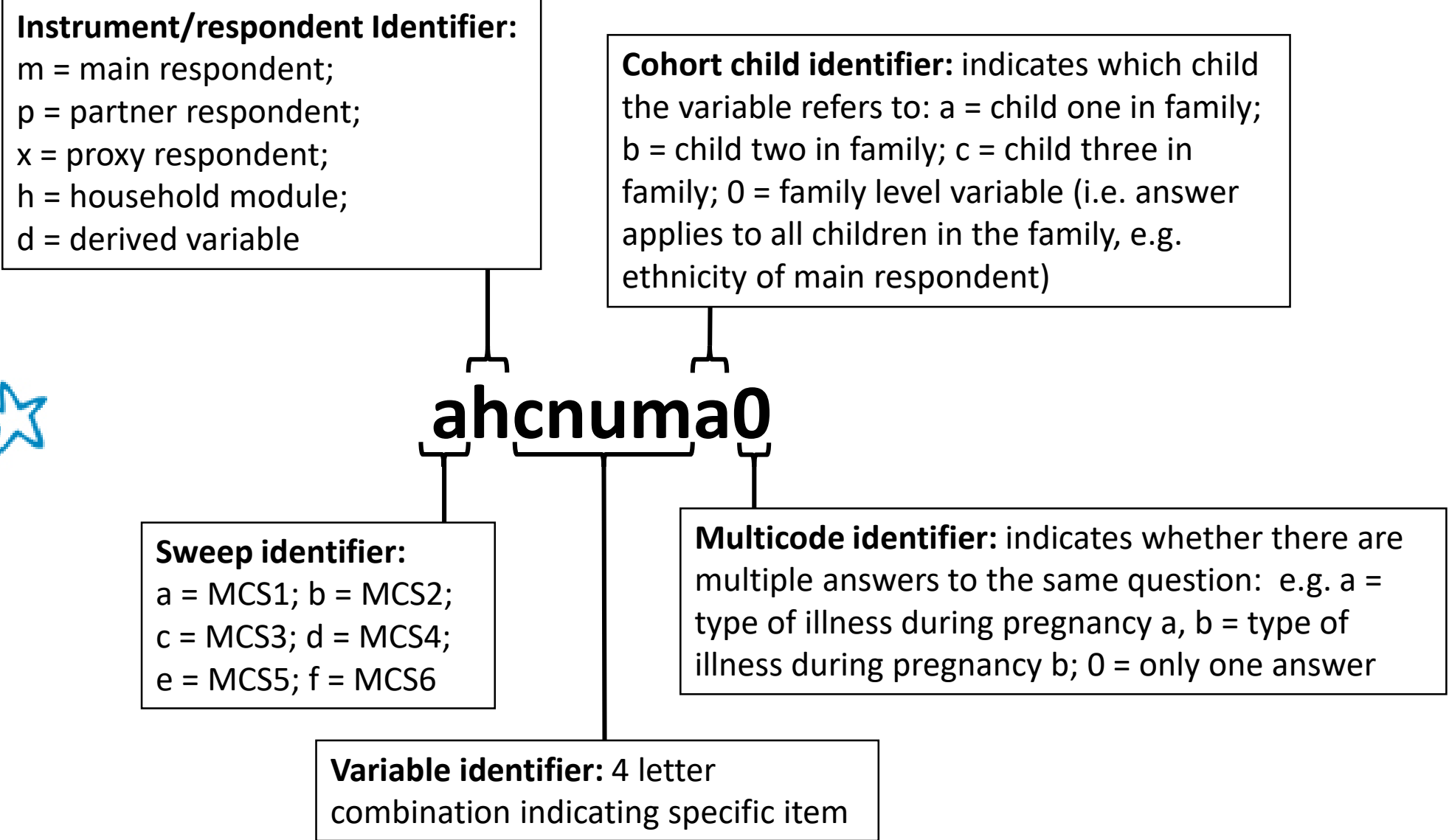
- Longitudinal study of around 19,000 children born in the UK between September 2000 and January 2002.
- To date, 7 sweeps of data have been collected
 - Ages: 9 months, 3, 5, 7, 11, 14, and 17 years
- Data are hosted by the Centre for Longitudinal studies and available to download for free from the UK Data Service



Packages

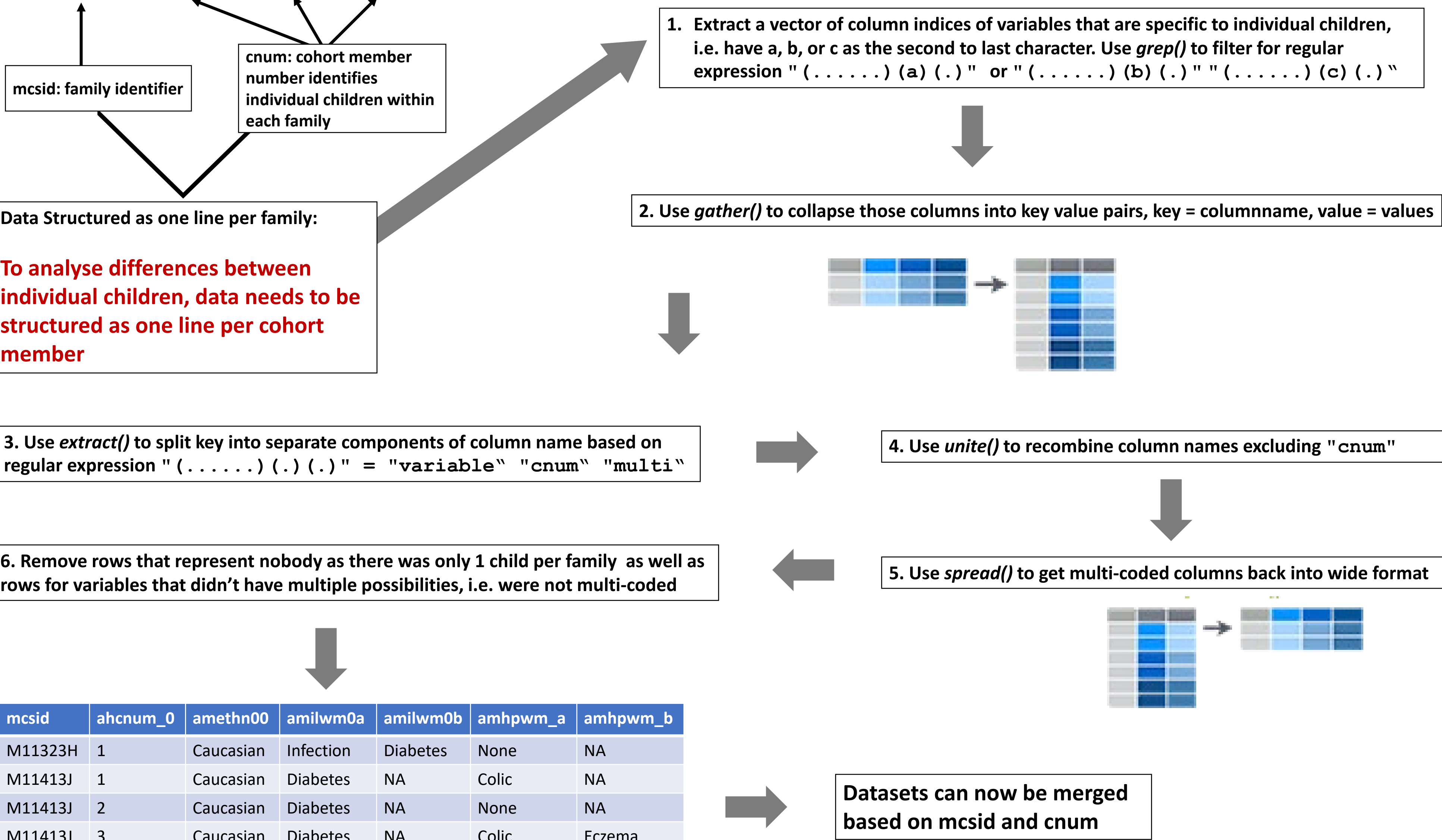
- haven*
 - to read spss files into R
- tidyverse/dplyr*
 - to restructure

Data Structure



Restructuring Data in R

mcsid	ahcnuma0	ahcnumb0	ahcnumc0	amethn00	amilwm0a	amilwm0b	amhpmwaa	amhpmwab	amhpmwba	amhpmwbb	amhpmwca	amhpmwcb
M11323H	1	NA	NA	Caucasian	Infection	Diabetes	None	NA	NA	NA	NA	NA
M11413J	1	2	3	Caucasian	Diabetes	NA	Colic	NA	None	NA	Colic	Eczema
M23453R	1	2	NA	Caucasian	None	NA	Eczema	Cold	Jaundice	NA	NA	NA
M21153S	1	NA	NA	Asian	Anaemia	NA	Colds	NA	NA	NA	NA	NA



mcsid	ahcnum_0	amethn00	amilwm0a	amilwm0b	amhpmw_a	amhpmw_b
M11323H	1	Caucasian	Infection	Diabetes	None	NA
M11413J	1	Caucasian	Diabetes	NA	Colic	NA
M11413J	2	Caucasian	Diabetes	NA	None	NA
M11413J	3	Caucasian	Diabetes	NA	Colic	Eczema
M23453R	1	Caucasian	None	NA	Eczema	Cold
M23453R	2	Caucasian	None	NA	Jaundice	NA
M21153S	1	Asian	Anaemia	NA	Colds	NA

Acknowledgements

Many thanks to the University of Edinburgh, Economic and Social Research Council and Scottish Graduate School for Social Sciences and for funding our work.

Access our Code



Notes on Analysis

- Analyses need to correct for:
 - Sampling Weights and Non-Response Bias
 - Finite Population Correction Factor
 - Clustering
 - Stratification
- survey package:**
 - descriptive statistics
 - linear regression
- lavaan.survey package:**
 - Structural equation modelling
 - Growth curve modelling
 - Factor analysis