CLOSER Work Package 1:

Harmonised Height, Weight and BMI User Guide

Prepared by: Rebecca Hardy, Jon Johnson, Alison Park (UCL)

July 2016

Contents

Preface

CLOSER (Cohort & Longitudinal Studies Enhancement Resources) aims to maximise the use, value and impact of longitudinal studies, both at home and abroad. Bringing together eight leading studies, the British Library and the UK Data Service, CLOSER works to stimulate interdisciplinary research, develop shared resources, provide training, and share expertise. In this way CLOSER is helping to build the body of knowledge on how life in the UK is changing – both across generations and in comparison to the rest of the world.

CLOSER's research includes a number of work packages focused on retrospective harmonisation, their aim being to make the data from different longitudinal studies more comparable in order to find out how life in the UK is changing from generation to generation. This documentation describes a dataset produced as part of the first CLOSER harmonisation work package, about weight and height.

Acknowledgements

CLOSER would like to thank the studies for providing the data and the participants of the 5 studies for taking part over many years to make this research possible.

Dataset Production

The dataset was constructed by Rebecca Hardy and Will Johnson both at UCL.

Introduction

This document explains the steps taken to harmonise anthropometric data across the UK birth cohort studies as part of CLOSER Work Package 1. This work focused on weight and height, which are each important indicators of lifelong health. The former reflects more acute changes in the nutritional environment and disease status e.g., weaning and infection, while the latter reflects more chronic exposures e.g., war and poverty (Gunnell (2002). Importantly, measurement of weight and height allows the computation of body mass index (BMI, kg/m2).

These datasets formed the basis of articles by Johnson et al (2015) and Bann et al (2017).

Studies included

1946 MRC National Survey of Health and Development (NSHD)
1958 National Child Development Study (NCDS)
1970 British Cohort Study (BCS70)
Avon Longitudinal Study of Parents and Children (ALSPAC)
Millennium Cohort Study (MCS)

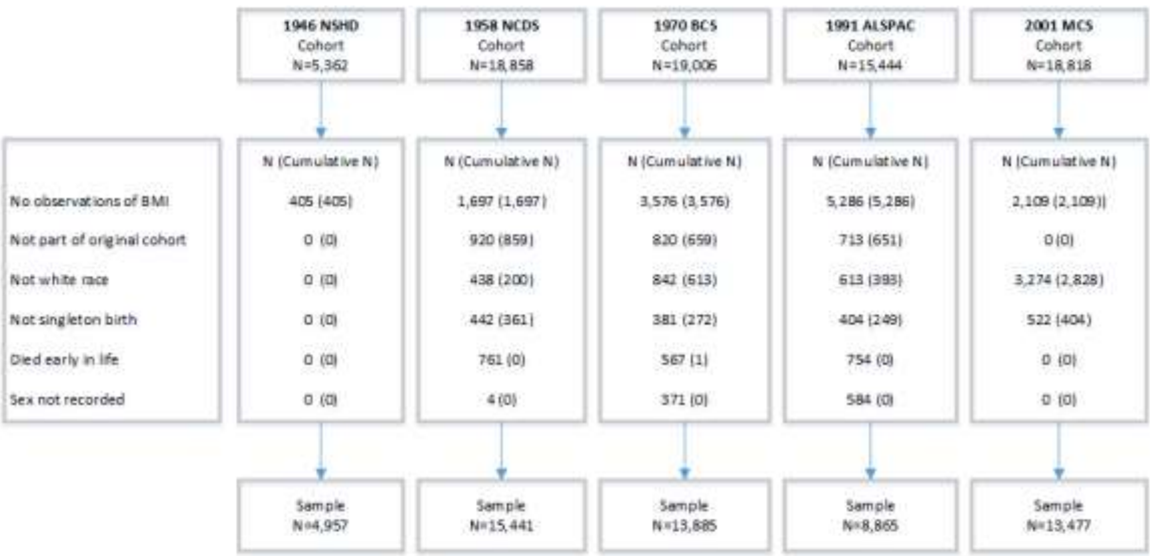
Methods

The BMI is an index of weight-for-height that works on the basis that weight increases proportionately to height squared, so dividing weight by height squared results in an index that is uncorrelated with height in a sample. This index was initially called the Quetelet Index (Quetelet, 1968) after the Belgian scientist Lambert Adolphe Quetelet (1796-1874) who first proposed the formula. By the mid-20th century, the index was used by some human biologists to adjust weight for height, and by life insurance companies to determine “desirable weight” ranges associated with the lowest mortality risks (Metropolitan Life Insurance Company, 1983). Alongside Quetelet’s Index others had, however, already suggested kg/m3, kg/m, kg/m-0.33.4 (Keys et al, 1972). In a review of the published literature in 1972, Ancel Keys (1904-2004) and colleagues set out to decide which was the best index based on 1) the strongest correlation with the numerator weight and with measurements of adiposity (as the impacts of high weight on disease risk were already known to be due to high body fat) and 2) the weakest correlation with the denominator height. Quetelet’s Index was found to be the best “obesity index” and a change in name to the BMI was proposed.

Despite BMI never being intended to assess adiposity i.e., thinness, normal weight, overweight, obese at the individual level i.e., it was developed to make weight independent of height at the population level, it is the most commonly used adiposity measure in clinical settings. Standard cut-offs of 25 and 30 kg/m2 are used to classify overweight and obesity in adulthood, respectively, and various age-specific cut-offs are available during childhood when BMI shows a normal pattern of age-related change (Cole, 2012). The BMI and it associated cut-offs have also become a central part of epidemiological research, namely because it is implicated in nearly all disease processes, either due to correlation with adiposity or due to individual values being dependent on other complicit factors e.g., height and fat-free mass (Wells, 2014). As such, it is unsurprising that the serial assessment of BMI has been a core part of the UK birth cohort studies measurement programmes.

Sample Selection

The sample used in these datasets were drawn using the following strategy. NB. The numbers for BCS70 vary slightly from that used in Johnson et al (2015) because there were a number of duplicate cases in the original dataset which were corrected.



Data Cleaning

Height & Weight

Details of the main differences in measurement protocols for the weight and height data (both between studies, and within the same study over time) are provided in Appendix 1.

Measured data at age 16 years in BCS70 were augmented with 2,353 self-reported weights and 2,309 self-reported heights at the same age to maximise the amount of available information. Further, measured data at age 44 years in the 1958 NCDS were augmented with 12 observations of self-reported weight greater than 150 kg. This was done in an attempt to retrieve information from the upper end of the distribution that appeared to have been removed by the employment of a cut-off during data entry or cleaning. Similar situations were found for weight at age 7 years in the 1958 NCDS, and weight at age 10 years and height at age 26 years in BCS70, but in these instances it was not possible to retrieve any data. Height was only reported at age 50 years in the 1958 NCDS if it had not been measured at the previous sweep at age 44 years; 9,063 missing observations of height at age 50 years were filled in with observations of height from the sweep at age 44 years. The same strategy had already been applied to study derived variables of height at ages 34 (filled in using age 30 year data) and 42 years (filled in using age 30 or 34 year data) in BCS70. Further details are in Appendix 2.

Date of interview / Assessment

For sweeps that were missing a date or some component of a date variable (i.e., day or month or year), day, month, and/ or year was assigned to the whole cohort; day was taken to be 15 in all studies, month to be 3 in the 1946 NSHD and 1958 NCDS and 4 in the 1970 BCS, and year to be that in which the sweep took place for the 1946 NSHD, 1958 NCDS, and 1970 BCS.

Age of participants

Where variables of decimal age were not available, they were computed from existing age variables or as the difference between date of birth and date of assessment. Participants who were still missing decimal age were assigned the mean value for that cohort at that sweep.

Precision of measurement

This indicates the precision (in kg or m) of the weight and height measurements e.g., weight was recorded to the last 0.1 kg. Further details are in Appendix 1.

Coding of missing on weight

Weight data was recoded to missing if a woman was pregnant at measurement. In some instances, no data on whether or not a woman was pregnant at a given sweep were recorded. Where it was possible to identify weight measurements taken while a woman was pregnant, these were recoded as missing (1946 NSHD: 257 observations; 1958 NCDS: 684 observations; 1970 BCS: 110 observations).

Height and Weight cut-offs

Observations were recoded as missing if the following criteria where met:

a. biologically implausible values using sensible yet arbitrary cut-offs (e.g., weight > 250 kg and height > 3 m) and;
b. inspection of a connected scatter plot of serial weight or height against age for persons with a measurement or change in measurement between two consecutive ages greater than five standard deviations from the sex and study stratified mean.

The total number of weight observations recoded as missing in the 1946 NSHD, 1958 NCDS, 1970 BCS, 1991 ALSPAC, and 2001 MCS were 3, 371, 50, 10, and 90, respectively. For height, these numbers were 15, 296, 100, 24, and 16.

Limitations

The strength of these data is the serial measurement of BMI in different nationally-representative birth year cohorts. The serial data allows investigation of age effects i.e., how BMI changes as a person ages; while comparison across studies allows investigation of cohort effects i.e., how BMI is different in someone born in the 1940's compared to someone born in the 1970's, at comparable ages; and period effects i.e., how some external event occurring at some point in time, such as a war or economic crisis, impacts on BMI in people of different ages (Bell, 2013).

.

Harmonised Variables Description and Source

Variable name:	wt				
Variable Description:	Harmonised Weight				
Source Variables					
Age	**NSHD**	**NCDS**	**BCS70**	**ALSPAC**	**MCS**
0	mbwt	n574	a0278		
1					amwtkga0 amwtlba0 amwtoua0
2	wt48				
3					bywtkm00 bywtgm00 bywtcm00
4	wt50				
5					cywtcm00 cywtgm00
6	wt52				
7	wt53	n337		f7ms026	dcwtcm00
8				f8lf021	
9				f9ms026	
10			meb19_1	fdms026	
11	wt57	n1515		fems026	ecwtcma0
14					
15	wt61				
16		n1953	rd4_1		
18				fjmr022	
20	wt66				
23		n5755 n5757			
26	wt72		b960439 b960441 b960443		
30			wtkilos2 wtstone2 wtpound2		
33		n504734			
34			b7wtkis2 b7wtste2 b7wtpod2		
36	wt82				
42		wtkilos2 wtstone2 wtpound2	wtkilos wtstones wtpounds		
43	wt89				
44		wtres			
50		n8wtkis2 n8wtste2 n8wtpod2			
53	wt99				
63	wtn09				

Variable name:	ht				
Variable Description:	Harmonised Height				
Source Variables					
Age	**NSHD**	**NCDS**	**BCS70**	**ALSPAC**	**MCS**
0					
1					
2	ht48				
3					byhtcm00 byhtmm00
4	ht50				
5			f102		cyhtcm00 cybkcm00
6	ht52				
7	ht53	n332 n334		f7ms010	dchtcm00
8				f8lf020	
9				f9ms010	
10			meb17	fdms010	
11	ht57	n1510 n1511		fems010	echtcma0
14					
15	ht61				
16		n1949	rd2_1		
18					
20	ht66				
23		n5752 n5753		fjmr020	
26	ht72		b960433 b960434 b960436 b960437		
30			htmetre2 htcms2 htfeet2 htinche2		
33		n504731			
34			b7htmees b7htcms b7htfeet b7htines		
36	ht82				
42		htmetre2 htcms2 htfeet2 htinche2	htmetres htcms htfeet htinches		
43	ht89				
44		htres			
50		n8htmees n8htcms n8htfeet n8htines			
53	ht99				
63	htn09				

Datasets

Study	No. of Cases	Dataset Name	Data Licencing
NSHD	4,957	nshd_closer_wp1	Special Licence
ALSPAC	8,665	alspac_closer_wp1	Special Licence
NCDS	15,441	ncds_bcs70_mcs_closer_wp1	End User Licence
BCS70	13,885		
MCS	13,477		

Licencing

All datasets are available from the UK Data Service (UKDS) at https://www.ukdataservice.ac.uk/get-data/key-data/cohort-and-longitudinal-studies. All users of the data need to be registered with the UKDS (details of how to do this are available at https://www.ukdataservice.ac.uk/get-data/how-to-access/registration. Data under the End User Licence can be downloaded once the access conditions have been ticked. Data under the Special Licence will need to request permission and complete a form. Once that has been accepted the data will be available to download.

Linkage to other data from a study

The datasets provided above will only be linkable to other data released by CLOSER under the CLOSERID (see below). If you wish to link a study to other data not released by CLOSER, data on the relevant study identifier from NCDS, BCS70 and MCS will be available from the UKDS at:

- NCDS - https://discover.ukdataservice.ac.uk/series/?sn=2000032
- BCS70 - https://discover.ukdataservice.ac.uk/series/?sn=200001
- MCS - https://discover.ukdataservice.ac.uk/series/?sn=2000031

If you wish to link further data to NSHD: contact them at mrclha.swiftinfo@ucl.ac.uk
If you wish to link further data to ALSPAC: contact them at https://proposals.epi.bristol.ac.uk/

Dataset structure

All datasets have the same structure

Variable name	Label	Format	Values
closerid	Participant identifier	String (8)	
stid	Study	integer	1 – nshd, 2 – ncds, 3 – bcs70, 4 – alspac, 5 - mcs
visitage	Age at visit/interview	decimal	
sex	Sex of participant	Integer	1 – male, 2 – female
xage	Actual age at measurement	decimal	
wt	Harmonised Weight	decimal	
wtself	Weight, measured or self-report	integer	1 – measured, 2 – selfreport
wtimp	Weight Imperial or metric	integer	1 – metric, 2 – imperial
wtpre	Weight (precision)	decimal	
ht	Harmonised Height	decimal	
htself	Height, measured or self-report	integer	1 – measured, 2 – selfreport
htimp	Height Imperial or metric	integer	1 – metric, 2 – imperial
htpre	Height (precision)	decimal	
bmi	Body Mass Index (kg/m2)	decimal	

Appendix 1

Main differences in measurement protocols for the weight and height data used

	Sweep Target age (date)	Assessment type	System of measurement	Precision of weight measurement	Precision of height measurement
1946 NSHD	2 (1948)	Measured (health visitor)	Imperial	0.028 kg	0.025 m
	4 (1950)	Measured (health visitor)	Imperial	0.113 kg	0.025 m
	6 (1952)	Measured (school doctor)	Imperial	0.113 kg	0.006 m
	7 (1953)	Measured (school doctor)	Imperial	0.028 kg	0.006 m
	11 (1957)	Measured (school doctor)	Imperial	0.028 kg	0.006 m
	15 (1961)	Measured (school doctor)	Imperial	0.028 kg	0.006 m
	20 (1966)	Self-reported (postal questionnaire)	Imperial	0.454 kg	0.025 m
	26 (1972)	Self-reported (administered questionnaire)	Imperial	0.454 kg	0.025 m
	36 (1982)	Measured (trained nurse)	Metric	0.5 kg	0.005 m
	43 (1989)	Measured (trained nurse)	Metric	0.5 kg	0.001 m
	53 (1999)	Measured (trained nurse)	Metric	0.1 kg	0.005 m
	60-64 (2006-2010)	Measured (trained nurse)	Metric	0.1 kg	0.001 m

	Sweep Target age (date)	Assessment type	System of measurement	Precision of weight measurement	Precision of height measurement
1958 NCDS	7 (1965)	Measured (medical officer)	Metric or imperial	0.454 kg	0.01 m
	11 (1969)	Measured (medical officer)	Imperial	0.454 kg	0.006 m
	16 (1974)	Measured (medical officer)	Metric or imperial	0.01 to 0.454 kg	0.006 to 0.01 m
	23 (1981)	Self-reported (administered questionnaire)	Imperial	0.454 kg	0.025 m
	33 (1991)	Measured (trained interviewer)	Metric	0.1 kg	0.01 m
	42 (2000)	Self-reported (CAPI)	Metric or imperial	0.454 to 1 kg	0.01 to 0.025 m
	44 (2002)	Measured (trained nurse) or self-reported (CAPI)	Metric or imperial	0.1 to 0.454 kg	0.001 m
	50 (2008)	Self-reported (CAPI)	Metric or imperial	0.454 to 1 kg	0.01 to 0.025 m

1970 BCS	10 (1980)	Measured (medical officer)	Metric or imperial	0.028 to 0.1 kg	0.001 to 0.006 m
	16 (1986)	Measured (medical officer) or self-reported (questionnaire)	Metric or imperial	0.028 to 0.1 kg	0.005 to 0.006 m
	26 (1996)	Self-reported (postal questionnaire)	Metric or imperial	0.454 to 1 kg	0.01 to 0.025 m
	30 (2000)	Self-reported (CAPI)	Metric or imperial	0.454 to 1 kg	0.01 to 0.025 m
	34 (2004)	Self-reported (CAPI)	Metric or imperial	0.454 to 1 kg	0.01 to 0.025 m
	42 (2012)	Self-reported (CAPI)	Metric or imperial	0.454 to 1 kg	0.01 m

1991 ALSPAC	7 (1998)	Measured (anthropometrist)	Metric	0.05 kg	0.001 m
	8 (1999)	Measured (anthropometrist)	Metric	0.05 kg	0.001 m
	9 (2000)	Measured (anthropometrist)	Metric	0.05 kg	0.001 m
	10 (2001)	Measured (anthropometrist)	Metric	0.05 kg	0.001 m
	11 (2002)	Measured (anthropometrist)	Metric	0.05 kg	0.001 m
	13 (2004)	Measured (anthropometrist)	Metric	0.1 kg	0.001 m
	14 (2005)	Measured (anthropometrist)	Metric	0.1 kg	0.001 m
	15 (2006)	Measured (anthropometrist)	Metric	0.1 kg	0.001 m
	18 (2009)	Measured (anthropometrist)	Metric	0.05 kg	0.001 m
2001 MCS	3 (2004)	Measured (trained interviewer)	Metric	0.001 kg	0.001 m
	5 (2006)	Measured (trained interviewer)	Metric	0.1 kg	0.001 m
	7 (2008)	Measured (trained interviewer)	Metric	0.1 kg	0.001 m
	11 (2012)	Measured (trained interviewer)	Metric	0.1 kg	0.001 m

CAPI: Computer-Assisted Personal Interviewing, NSHD: Medical Research Council National Survey of Health and Development, NCDS National Child Development Study, BCS: British Cohort Study, ALSPAC: Avon Longitudinal Study of Parents and Children, MCS: Millennium Cohort Study

Appendix 2

Height and Weight Harmonisation

1	Weights and heights were converted to kilograms and metres, respectively, as necessary.
2	Measured data at age 16 years in the 1970 BCS were augmented with 2,353 self-reported weights and 2,309 self-reported heights at the same age to maximise the amount of available information.
3	Further, measured data at age 44 years in the 1958 NCDS were augmented with 12 observations of self-reported weight greater than 150 kg. This was done in an attempt to retrieve information from the upper end of the distribution that appeared to have been removed by the employment of a cut-off during data entry or cleaning. Similar situations were found for weight at age 7 years in the 1958 NCDS, and weight at age 10 years and height at age 26 years in the 1970 BCS, but in these instances it was not possible to retrieve any data.
4	Height was only reported at age 50 years in the 1958 NCDS if it had not been measured at the previous sweep at age 44 years; 9,063 missing observations of height at age 50 years were filled in with observations of height from the sweep at age 44 years. The same strategy had already been applied to study derived variables of height at ages 34 (filled in using age 30 year data) and 42 years (filled in using age 30 or 34 year data) in the 1970 BCS.
5	Where variables of decimal age at assessment were not available, they were computed from existing age variables or as the difference between date of birth and date of assessment.
6	For sweeps that were missing a date or some component of a date variable (i.e., day or month or year), day, month, and/ or year was assigned to the whole cohort; day was taken to be 15 in all studies, month to be 3 in the 1946 NSHD and 1958 NCDS and 4 in the 1970 BCS, and year to be that in which the sweep took place for the 1946 NSHD, 1958 NCDS, and 1970 BCS.
7	Participants who were still missing decimal age were assigned the mean value for that cohort at that sweep.
8	In some instances, no data on whether or not a woman was pregnant at a given sweep were recorded. Where it was possible to identify measurements taken while a woman was pregnant, these were excluded (1946 NSHD: 257 observations; 1958 NCDS: 684 observations; 1970 BCS: 110 observations).
9	A standardised data cleaning protocol was applied. This involved removal of biologically implausible values using sensible yet arbitrary cut-offs (e.g., weight > 250 kg and height > 3 m) and inspection of a connected scatter plot of serial weight or height against age (i.e., a trajectory) for persons with a measurement or change in measurement between two consecutive ages greater than five standard deviations from the sex and study stratified mean. The total number of weight observations excluded by this cleaning process in the 1946 NSHD, 1958 NCDS, 1970 BCS, 1991 ALSPAC, and 2001 MCS were 3, 371, 50, 10, and 90, respectively. For height, these numbers were 15, 296, 100, 24, and 16.

NSHD: Medical Research Council National Survey of Health and Development, NCDS National Child Development Study, BCS: British Cohort Study, ALSPAC: Avon Longitudinal Study of Parents and Children, MCS: Millennium Cohort Study

Appendix 3

Derivation Code

These are supplied as separate stata .do files (one per study) alongside the documentation.

alspac code.do
bcs code.do
mcs code.do
ncds code.do
nshd code.do

Source Files

Files from which the derivations were made are:

Study	Age	Data Source
NSHD	All	Provided by study
NCDS	7	http://doi.org/10.5255/UKDA-SN-5565-2
	11	
	16	
	23	http://doi.org/10.5255/UKDA-SN-5566-1
	33	http://doi.org/10.5255/UKDA-SN-5567-1
	42	http://doi.org/10.5255/UKDA-SN-5578-1
	44	http://doi.org/10.5255/UKDA-SN-5594-2
	50	http://doi.org/10.5255/UKDA-SN-6137-2
BCS70	5	http://doi.org/10.5255/UKDA-SN-2699-4
	10	http://doi.org/10.5255/UKDA-SN-3723-7
	16	http://doi.org/10.5255/UKDA-SN-3535-4
	26	http://doi.org/10.5255/UKDA-SN-3833-3
	30	http://doi.org/10.5255/UKDA-SN-5558-3
	34	http://doi.org/10.5255/UKDA-SN-5585-3
	42	http://doi.org/10.5255/UKDA-SN-7473-2
NCDS	7	http://doi.org/10.5255/UKDA-SN-5565-2
	11	
	16	
	23	http://doi.org/10.5255/UKDA-SN-5566-1
	33	http://doi.org/10.5255/UKDA-SN-5567-1
	42	http://doi.org/10.5255/UKDA-SN-5578-1
	44	http://doi.org/10.5255/UKDA-SN-5594-2
	50	http://doi.org/10.5255/UKDA-SN-6137-2
ALSPAC	All	Provided by study
MCS	1	http://doi.org/10.5255/UKDA-SN-4683-4
	3	http://doi.org/10.5255/UKDA-SN-5350-4
	5	http://doi.org/10.5255/UKDA-SN-5795-4
	7	http://doi.org/10.5255/UKDA-SN-6411-7
	11	http://doi.org/10.5255/UKDA-SN-7464-4

References

Bann D, Johnson W, Li L, Kuh D, Hardy R (2017) Socioeconomic Inequalities in Body Mass Index across Adulthood: Coordinated Analyses of Individual Participant Data from Three British Birth Cohort Studies Initiated in 1946, 1958 and 1970. PLoS Med 14(1): e1002214. doi:10.1371/journal. pmed.1002214

Bell A, Jones K. (2013) The impossibility of separating age, period and cohort effects. Soc Sci Med. 93:163-165.

Cole TJ, Lobstein T. (2012) Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes. 7:284-294.

Gunnell D. (2002) Can adult anthropometry be used as a 'biomarker' for prenatal and childhood exposures? Int J Epidemiol. 31:390-394.

Johnson W, Li L, Kuh D, Hardy R. (2015) How Has the Age-Related Process of Overweight or Obesity Development Changed over Time? Co-ordinated Analyses of Individual Participant Data from Five United Kingdom Birth Cohorts. PLoS Med 12(5): e1001828. doi:10.1371/journal.pmed.1001828

Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. (1972) Indices of relative weight and obesity. J Chronic Dis. 25:329-343.

Metropolitan Life Insurance Company. Metropolitan height and weight tables. (1983) Stat Bull Metropol Life Insur Co. 64:2-9.

Quetelet A. (1968) A treatise on man and the development of his faculties. Reprinted New York, USA: Burt Franklin. 1842.

Wells JC. (2014) Commentary: The paradox of body mass index in obesity assessment: not a good index of adiposity, but not a bad index of cardio-metabolic risk. Int J Epidemiol. 43:672-674.