Completeness and representativeness of BMI in children's electronic general practice records: linked cross-sectional study in an ethnically-diverse urban population in the United Kingdom

Nicola Firman[1], John Robson[1], Zaheer Ahmed[1], Kambiz Boomla[1], Carol Dezateux[1]

[1] Centre for Clinical Effectiveness and Health Data Science, Institute of Population Health Sciences, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Abernethy Building, 4 Newark Street, London, United Kingdom, E1 2AT

Contact information: Nicola Firman; Nicola.firman@qmul.ac.uk; Centre for Clinical Effectiveness and Health Data Science, Institute of Population Health Sciences, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Abernethy Building, 4 Newark Street, London, United Kingdom, E1 2AT

Table S1 – Characteristics of linked and unlinked sample

	Linked (*n*=53,695)			Not linked (*n*=2,804*)*		
	n	%	95% CI[1]	n	%	95% CI[1]
Local authority[2]						
City and Hackney	18538	34.5	34.1,34.9	1550	55.3	53.4,57.1
Newham	24535	45.7	45.2,46.1	992	35.4	33.6,67.2
Tower Hamlets	10622	19.8	19.4,20.1	262	9.3	8.3,10.5
School year						
2013/14	4399	8.2	8.0,8.4	374	13.3	12.1,14.6
2014/15	12736	23.7	23.4,24.1	713	25.4	23.9,27.1
2015/16	17472	32.5	32.1,32.9	852	30.4	28.7,32.1
2016/17	19088	35.6	35.1,36.0	865	30.9	29.2,32.6
Year group						
Age 5	28330	52.8	52.3,53.2	1509	53.8	52.0,55.7,
Age 11	25365	47.2	46.8,47.7	1295	46.2	44.3,48.0
Sex						
Male	27289	50.8	50.4,51.2	1446	51.6	49.7,53.4
Female	26406	49.2	48.8,49.6	1358	48.4	46.6,50.3
Ethnic background[3]						
White	11695	21.8	21.4,22.1	580	20.7	19.2,22.2
Mixed and Other	10386	19.3	19.0,19.7	764	27.3	25.6,28.9
South Asian	18728	34.9	34.5,35.3	500	17.8	16.5,19.3
Black	11072	20.6	20.3,21.0	596	21.2	19.8,22.8
Missing	1814	3.4	3.2,3.5	364	13.0	11.8,14.3
IMD quintile[4]						
1 - most deprived	34746	64.7	64.3,65.1	1526	54.4	52.6,56.3
2	16817	31.3	30.9,31.7	1024	36.5	34.8,38.3
3	1590	3.0	2.8,3.1	135	4.8	4.1,5.7
4	392	0.7	0.7,0.8	84	3.0	2.4,3.7
5 - least deprived	75	0.2	0.1,0.2	34	1.2	0.9,1.7
Missing	75	0.1	0.1,0.2	1	0.01	0.01,0.3
NCMP weight status[5]						
Underweight	1679	3.1	3.0,3.3	56	2.0	1.5,2.6
Healthy weight	38272	71.3	70.9,71.7	2075	74.0	72.3,75.6
Overweight	6981	13.0	12.7,13.3	339	12.1	10.9,13.3
Obese	6763	12.6	12.3,12.9	334	11.9	10.8,13.2

[1] 95% confidence interval. [2] Local authority which collected the National Child Measurement Programme (NCMP) data. [3] Ethnic background as recorded in the NCMP and supplemented with GP ethnicity for those with missing NCMP ethnic group. [4] Index of Multiple Deprivation (IMD) quintile based on NCMP-recorded child's home address postcode. [5] Weight status based on NCMP-recorded BMI categorised according to UK1990 clinical reference standard: "underweight" (BMI<2nd centile), "healthy weight" (≥2nd to <91st centile), "overweight" (≥91st to <98th centile) or "obese" (≥98th centile).

Table S2 – Weight status distribution according to International Obesity Task Force (IOTF), World Health Organization (WHO) and UK1990 reference standards

	Age 5			Age 11			All		
	n	%	95% CI[1]	n	%	95% CI[1]	n	%	95% CI[1]
IOTF[2]									
Underweight	2538	8.9	8.6,9.3	1864	7.4	7.0,7.7	4402	8.2	8.0,8.4
Healthy weight	19710	69.6	69.0,70.1	14734	58.1	57.5,58.7	34444	64.1	63.7,64.6
Overweight	4098	14.5	14.1,14.9	6144	24.2	23.7,24.8	10242	19.1	18.7,19.4
Obese	1984	7.0	6.7,7.3	2623	10.3	10.0,10.7	4607	8.6	8.3,8.8
WHO[3]									
Underweight	594	2.1	1.9,2.3	521	2.0	1.9,2.2	1115	2.1	2.0,2.2
Healthy weight	20435	72.1	71.6,72.7	13891	54.8	54.2,55.4	34326	63.9	63.5,64.3
Overweight	4810	17.0	16.5,17.4	6233	24.6	24.0,25.1	11043	20.6	20.2,20.9
Obese	2491	8.8	8.5,9.1	4720	18.6	18.1,19.1	7211	13.4	13.1,13.7
UK1990 clinical[4]									
Underweight	537	1.9	1.7,2.1	479	1.9	1.7,2.1	1016	1.9	1.8,2.0
Healthy weight	22743	80.3	79.8,80.7	16192	63.8	63.2,64.4	38935	72.5	72.1,72.9
Overweight	2718	9.6	9.3,9.9	4263	16.8	16.4,17.3	6981	13.0	12.7,13.3
Obese	2332	8.2	7.9,8.6	4431	17.5	17.0,17.9	6763	12.6	12.3,12.9
UK1990 population[5]									
Underweight	537	1.9	1.7,2.1	479	1.9	1.7,2.1	1016	1.9	1.8,2.0
Healthy weight	20741	73.2	72.7,73.7	14172	55.9	55.3,56.5	34913	65.0	64.6,65.4
Overweight	3393	12.0	11.6,12.4	3892	15.3	14.9,15.8	7285	13.6	13.3,13.9
Obese	3659	12.9	12.5,13.3	6822	26.9	26.4,27.4	10481	19.5	19.2,19.9

[1] 95% confidence interval. [2] National Child Measurement Programme (NCMP)-recorded BMI categorised according to International Obesity Task Force cut-offs. [3] NCMP-recorded BMI according to World Health Organization cut-offs: "underweight" (BMI z-score <-2), "healthy weight" (BMI z-score ≥-2 to <1.0), "overweight" (BMI z-score ≥1.0 to <2.0), "obese" (BMI z-score ≥2.0). [4] NCMP-recorded BMI categorised according to UK1990 clinical reference standard: "underweight" (BMI<2nd centile), "healthy weight" (≥2nd to <91st centile), "overweight" (≥91st to <98th centile) or "obese" (≥98th centile). [5] NCMP-recorded BMI categorised according to UK1990 population reference standard: "underweight" (BMI<2nd centile), "healthy weight" (≥2nd to <85th centile), "overweight" (≥85th to <95th centile) or "obese" (≥95th centile).

Table S3 – Univariable and adjusted odds of ever having a GP-BMI

	Age 5				Age 11			
	OR[1]	95% CI[2]	aOR[3]	95% CI[2]	OR[1]	95% CI[2]	aOR[3]	95% CI[2]
Local authority[4]								
City and Hackney (ref.[5])	1		1		1		1	
Newham	0.89	0.82,0.97	0.84	0.75,0.92	1.05	0.99,1.12	1.02	0.96,1.11
Tower Hamlets	0.94	0.84,1.04	0.84	0.72,0.93	0.95	0.88,1.03	0.81	0.73,0.89
Sex								
Male (ref.)	1		1		1		1	
Female	0.84	0.78,0.91	0.93	0.86,1.01	0.88	0.83,0.93	0.99	0.94,1.06
Ethnic background[6]								
White (ref.)	1		1		1		1	
Mixed and Other	1.08	0.95,1.23	1.09	0.96,1.24	0.97	0.88,1.06	0.92	0.83,1.01
South Asian	1.46	1.32,1.62	1.55	1.38,1.74	1.25	1.15,1.35	1.27	1.16,1.39
Black	1.17	1.03,1.31	1.09	0.96,1.24	0.99	0.91,1.09	0.94	0.85,1.03
IMD quintile[7]								
1 - most deprived (ref.)	1		1		1		1	
2	0.76	0.70,0.83	0.77	0.70,0.84	0.90	0.85,0.96	0.88	0.82,0.95
3	0.61	0.46,0.79	0.70	0.53,0.93	0.70	0.58,0.83	0.70	0.58,0.86
4	0.50	0.29,0.88	0.56	0.30,1.02	0.38	0.24,0.60	0.39	0.24,0.64
5 - least deprived	0.51	0.16,1.63	0.77	0.24,2.50	1	0.42,2.38	0.78	0.27,2.25
NCMP weight status[8]								
Underweight	1.89	1.50,2.37	1.71	1.34,2.19	1.85	1.53,2.24	1.99	1.62,2.44
Healthy weight (ref.)	1		1		1		1	
Overweight	1.09	0.96,1.25	1.03	0.90,1.18	1.06	0.98,1.15	1.01	0.93,1.10
Obese	1.53	1.35,1.73	1.45	1.27,1.65	1.54	1.44,1.66	1.50	1.39,1.62
Long-term condition[9]								
No (ref.)	1		1		1		1	
Yes	8.38	7.53,9.33	8.15	7.30,9.09	8.43	7.72,9.21	8.53	7.79,9.35

[1] Univariable odds ratio. [2] 95% confidence interval. [3] Adjusted odds ratio after mutual adjustment for local authority, sex, ethnic background, Index of Multiple Deprivation (IMD) quintile, National Child Measurement Programme (NCMP) weight status and presence of a long-term condition. [4] Local authority which collected the NCMP data. [5] Reference category. [6] Ethnic background as recorded in the NCMP and supplemented with GP ethnicity for those with missing NCMP ethnic group. [7] IMD quintile based on NCMP recorded child's home address postcode. [8] Weight status based on NCMP recorded BMI and categorised according to UK1990 clinical reference standard: "underweight" (BMI<2nd centile), "healthy weight" (≥2nd to <91st centile), "overweight" (≥91st to <98th centile) or "obese" (≥98th centile). [9] Long-term conditions included GP recorded diagnosis of: cystic fibrosis, type 1 diabetes, attention deficit hyperactivity disorder, autism or learning disability or GP prescriptions for: epilepsy, attention deficit hyperactivity disorder, or thyroid disease. Asthma was included as a long-term condition if both GP diagnosis and prescription were present.

Figure S1 – Flow diagram explaining identification of height and weight measurements to derive GP-BMI

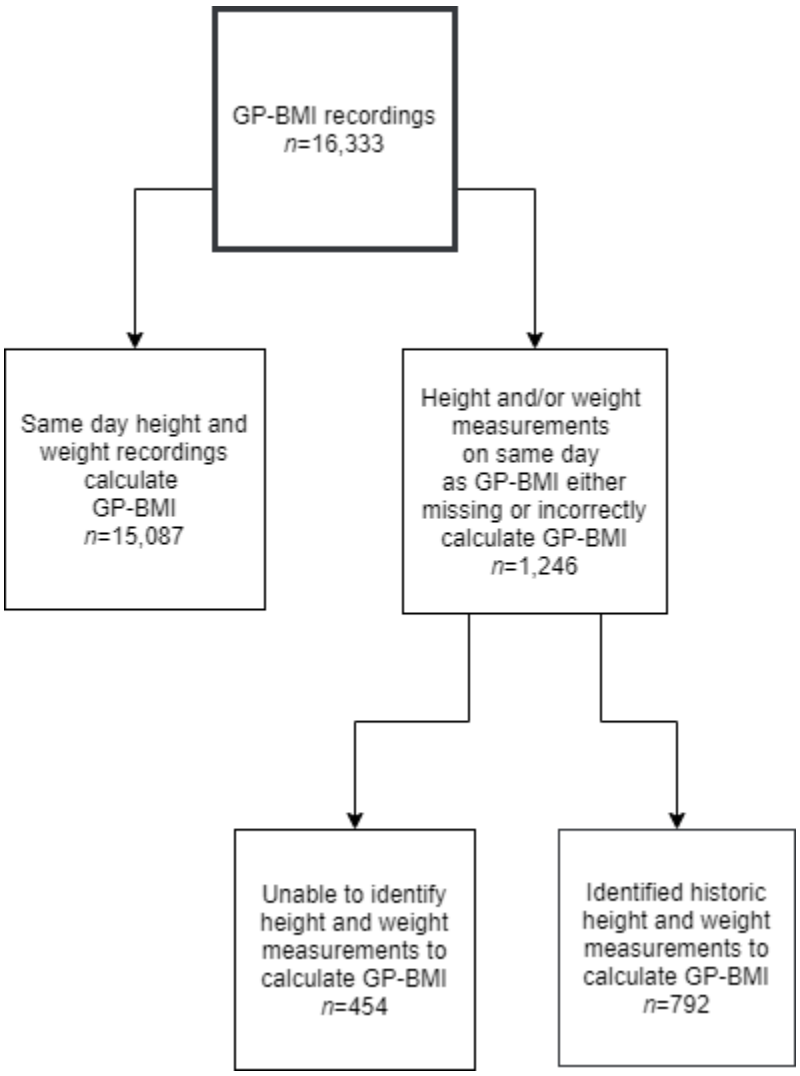

Table S4 – Prevalence of specified weight status[1] by data source

	NCMP[2]			GP-EHR		
	n	**%**	**95% CI[3]**	***n***	**%**	**95% CI[3]**
Age 5						
Underweight	537	1.9	1.7,2.1	103	8.8	7.3,10.5
Healthy weight	20741	73.2	72.7,73.7	765	65.1	62.3,67.8
Overweight	3393	12.0	11.6,12.3	93	7.9	6.5,9.6
Obese	3659	12.9	12.5,13.3	214	18.2	16.1,20.5
Age 11						
Underweight	479	1.9	1.7,2.1	72	3.5	2.7,4.3
Healthy weight	14172	55.9	55.3,56.5	979	46.8	44.7,49.0
Overweight	3892	15.3	14.9,15.8	289	13.8	12.4,15.4
Obese	6822	26.9	26.4,27.4	751	35.9	33.9,38.0

[1] Categorised according to UK1990 population reference standard: "underweight" (BMI<2nd centile), "healthy weight" (≥2nd to <85th centile), "overweight" (≥85th to <95th centile) or "obese" (≥95th centile). [2] National Child Measurement Programme. [3] 95% confidence interval. Median age in years (interquartile range) at NCMP-BMI measurement: 5.07 (4.81,5.33) and 10.93 (10.66,11.21). Median age in years at GP-BMI measurement: 5.03 (4.65,5.40) and 10.92 (10.48,11.34).

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Pages 1-2	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Pages 1-2 Page 2 Pages 1-2
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 3		
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 3		
Methods					
Study Design	4	Present key elements of study design early in the paper	Page 4		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 4		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources	n/a n/a	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.	Page 4

		and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	Page 4 n/a	RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	n/a Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Pages 4-7	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Pages 4-7
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Pages 4-7		
Bias	9	Describe any efforts to address potential sources of bias	Page 6		
Study size	10	Explain how the study size was arrived at	Pages 4 and 7		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Pages 4-7		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	Pages 6-7		

		(b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	n/a Pages 5-6 n/a n/a Pages 6-7 n/a		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Page 4 Pages 4-7
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Page 4
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage.	Page 7 Figure 1	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Pages 4-7

		(c) Consider use of a flow diagram	Figure 1		
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders	Page 7		
		(b) Indicate the number of participants with missing data for each variable of interest	Table 1		
		(c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	n/a		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time	n/a		
		Case-control study - Report numbers in each exposure category, or summary measures of exposure	n/a		
		Cross-sectional study - Report numbers of outcome events or summary measures	Pages 7-8		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table S3		
		(b) Report category boundaries when continuous variables were categorized	n/a		
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and	n/a		

		interactions, and sensitivity analyses			
Discussion					
Key results	18	Summarise key results with reference to study objectives	Page 9		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Pages 9-10	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Pages 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pages 12-13		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Pages 10-12		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 13		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 14

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.