Title: The association between childhood adiposity and appetite assessed using the Child Eating Behaviour Questionnaire and Baby Eating Behaviour Questionnaire: A systematic review and meta-analysis.

Authors: Alice Kininmonth[1], MSc*; Andrea Smith, PhD*[2]; Susan Carnell[3], PhD; Silje Steinsbekk[4], PhD; Alison Fildes[1], PhD*; and Clare Llewellyn[2], PhD*
**Joint senior author, *Joint first author*

[1] School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom.
[2] Obesity Group, Research Department of Behavioural Science and Health, University College London, London, United Kingdom.
[3] Department of Psychiatry and Behavioural Sciences, Johns Hopkins University School of Medicine, Baltimore, USA.
[4] Department of Psychology, Faculty of Social and Educational Sciences, Norwegian University of Science and Technology.

Corresponding author:
Dr Alison Fildes
School of Psychology,
Faculty of Medicine & Health
University of Leeds
Leeds, LS2 9JT
United Kingdom.
E-mail: A.Fildes@leeds.ac.uk

Supporting information

Table S1. Systematic search strategy

	Search strategy used in Ovid Medline based on PICO framework
1	exp CHILD/ or exp ADOLESCENT/ or exp CHILD, PRESCHOOL/ or CHILD/ or exp INFANT/
2	(child* or adolescen* or infant*).mp
3	(teenage* or young people or young person or young adult*).mp.
4	(schoolchildren or school children).mp.
5	(pediatr* or paediatr*).mp
6	(boys or girls or youth or youths).mp.
7	1 or 2 or 3 or 4 or 5 or 6
8	(Child* Eating Behavio?r Questionnaire or CEBQ).mp or appet* traits.mp or eating behaviours.mp or food fussiness.mp or food responsiveness.mp
9	(Emotional over-eating or emotional overeating or emotional eating or emotional over eating).mp
10	(enjoyment of food or desire to drink or satiety responsiveness or slowness in eating).mp
11	(Emotional under-eating or emotional undereating or emotional under eating)).mp.
12	8 or 9 or 10 or 11
13	(adipos* or (weight or weight status)).mp. or exp obesity/ or exp overweight/
14	BMI z-scores.mp. or BMI-for-age.mp. or weight-for-length percentiles.mp.
15	((weight-for-height percentiles or waist circumference) adj2 growth charts).mp. or skinfold thickness.mp. or Anthropometr*.mp. or
16	((weight or bmi or body mass index) adj2 (gain or loss or change or reduc*)).mp.
17	13 or 14 or 15 or 16
18	7 and 12 and 17
19	limit 18 to yr="2001 -Current"

**The syntax used in this search strategy was adjusted where necessary according to the requirements of each database.*

Table S2. Newcastle-Ottawa quality assessment scale for all included studies.

	Selection				Comparability		Outcome		Total Score
Cross-sectional CEBQ studies (n=43)	Representativeness of the sample	Sample size	Non-respondents	Ascertainment of exposure	Controls for most important confounder	Controls for additional confounders	Assessment of outcome	Statistical test	
Carnell & Wardle, 2008[§]	B*	A*	A*	B*	A*	A*	A**/D[‡]	A*	7/9[#]
Cao, 2012	C	B	A*	A*	A*	A*	A**	B	6
Bergmeier, 2014	B*	B	A*	B*	A*	A*	A**/D[‖]	A*	6/8[#]
Boswell, 2018[§]	A*	A*	A*	B*	A*	A*	D	A*	7
Braden, 2014†	C	B	B	B*	A*	A*	A**	A*	6
Brown, 2012	B*	B	A*	B*	A*	A*	D	B	5
Cross, 2014[§†]	C	B	A*	B*	A*	A*	A**	A*	7
Demir, 2017	A*	A*	C	B*	B	A*	D	A*	5
Domoff, 2015[§†]	C	B	B	A*	B	B	A**	A*	4
Emond, 2017[§†]	C	B	B	B*	A*	A*	A**	A*	6
Escobar, 2014[§†]	A*	B	C	B*	A*	A*	A**	A*	7
Frankel, 2014[§]	C	B	C	B*	A*	A*	A**	A*	6
Fuemmeler, 2013[§†]	C	B	B	B*	A*	A*	A**	A*	6
Gregory, 2010[§]	B*	B	A*	A*	A*	A*	D	A*	6
Hankey, 2016[§]	B*	A*	A*	B*	A*	A*	A**	A*	9
Hardman, 2016[§†]	B*	B	C	B*	B	A*	D	A*	4
Haycraft, 2011[§†]	B*	B	B	B*	A*	A*	D	A*	5
Hayes, 2016[§]	C	B	A*	B*	B	B	A**	A*	5
Jansen, 2012[§]	A*	B	B	A*	A*	A*	A**	A*	7
Koch, 2014[§]	A*	A*	A*	B*	B	B	A**	A*	7
Larsen, 2017[§]	B*	B	C	B*	A*	A*	A**	A*	7
Lipowska, 2018	A*	A*	A*	B*	A*	A*	A**	A*	9
Loh, 2013[§]	B*	A*	A*	A*	A*	A*	A**	A*	9
Lora, 2016[†]	B*	B	A*	A*	A*	A*	A**	A*	8
Mallan, 2013	B*	B	B	B*	A*	A*	A**/B[¶]	A*	7/5[#]
McCarthy, 2015[†]	A*	A*	A*	B*	A*	A*	A**	A*	9
McPhie, 2011[§]	B*	A*	B	B*	A*	A*	D	A*	6

Parkinson, 2010	A*	A*	A*	B*	A*	A*	A**	A*	9
Pesch, 2018	C	A*	A*	B*	A*	A*	A**	A*	8
Quah, 2017[§†]	A*	A*	A*	B*	A*	A*	A**	A*	9
Roach, 2017	C	A*	B	B*	A*	A*	A**	A*	7
Rudy, 2016[§]	C	A*	B	A*	A*	A*	A**	C	6
Sanchez, 2016[§†]	A*	A*	A*	B*	A*	A*	A**	A*	9
Sanlier, 2016	C	B	C	B*	A*	A*	A**	A*	6
Silva Garcia, 2016[§†]	C	B	B	B*	A*	A*	A**	A*	6
Sleddens, 2008[§]	B*	B	A*	B*	A*	A*	D	A*	7
Somaraki, 2018	C	A*	B	B*	A*	A*	D	A*	5
Soussignan, 2012[§†]	B	A*	A*	B*	A*	A*	D	A*	6
Svensson, 2011	B	A*	B	B*	A*	A*	B*	A*	6
Tay, 2016[§†]	A*	A*	A*	B*	A*	A*	A**	A*	9
Viana, 2008[§]	B*	B	C	B*	A*	A*	D	A*	5
Vollmer, 2015[§†]	B	A*	C	B*	A*	A*	A**	A*	7
Webber, 2009[§†]	B	A*	A*	B*	A*	A*	A**	A	7

	Selection				**Comparability**		**Outcome**		
Longitudinal CEBQ studies (n=12)	Representativeness of the sample	Sample size	Non-respondents	Ascertainment of exposure	Controls for most important confounder	Controls for additional confounders	Assessment of outcome	Statistical test	Total Score
Mallan, 2016	B*	B	B	B*	A*	A*	A**	A*	7
Mallan, 2014	C	B	A*	B*	A*	A*	A**	A*	7
McPhie, 2012	B*	B	C	B*	B	B	D	A*	3
Quah, 2015 [†]	B*	A*	B	B*	A*	A*	A**	A*	8
Steinsbekk, 2015	A*	A*	A*	B*	A*	A*	A**	A*	9
Steinsbekk, 2016	A*	A*	A*	B*	A*	A*	A**	A*	9
Derks, 2018	A*	A*	A*	B*	A*	A*	A**	A*	9
Steinsbekk, 2017	A*	A*	A*	B*	A*	A*	A**	A*	9
Bjorklund, 2018	A*	A*	A*	B*	A*	A*	A**	A*	9
Bergmeier, 2014	A*	A*	A*	B*	A*	A*	D	A*	7
Escobar, 2014 [†]	A*	A*	A*	B*	A*	A*	A**	A*	9
Parkinson, 2010	A*	A*	A*	B*	A*	A*	A**	A*	9
van Deutekom, 2016 [†]	A*	A*	A*	B*	A*	A*	B**	A*	9

	Selection				**Comparability**		**Outcome**		
BEBQ Studies (n=5)	Representativeness of the sample	Sample size	Non-respondents	Ascertainment of exposure	Controls for most important confounder	Controls for additional confounders	Assessment of outcome	Statistical test	Total Score
Mallan, 2014	B*	A*	A*	B*	B	B	B**	A*	7

Quah, 2015	B*	A*	B	B*	A*	A*	A**	A*	8
Shepard, 2015	B*	B	A*	B*	B	B	A**	A*	6
van Jaarsveld, 2015	B*	A*	A*	B*	A*	A*	D	A*	7
Patel, 2018	C	A*	A*	B*	A*	A*	A**	A*	8

† Indicates studies for which authors provided additional data.
‡ Weight outcome measured differently in sub cohorts. TEDS is parent reported BMI, and community sample is researcher-measured BMI.
§ Indicates studies included in the meta-analysis.
‖ Weight outcome measured differently in sub cohorts. Half the cohort provided parent reported BMI, and half via standardised weight measurement during home visit.
¶ Weight outcome measured differently in sub cohorts. Sample 1 provided researcher-measured weight, and Sample 2 & 3 provided mother-reported weight.
# Different values for Total Score indicate studies where quality of outcome assessment differed across sub cohorts, resulting in sub cohort specific total NOS scores.

Key to sub-component ratings (max 10 stars). A NOS score of >7 is considered a "good" study, and this was used as a cut-off to classify good study quality.

Selection (max 5 stars**)**
1) Representativeness of the exposed cohort
A*= Truly representative of the general population
B* = Somewhat representative of the general population
C = Selected group of users e.g. nurses, volunteers
2) Sample size
A* = Justified and satisfactory
B = Not justified
3) Non-respondents
A* = Comparability between respondents' and non-respondents' characteristics is established, and the response rate is satisfactory
B = The response rate is unsatisfactory, or the comparability between respondents and non-respondents is unsatisfactory.
C = No description of the response rate or the characteristics of the respondents and the non-respondents
4) Ascertainment of exposure (CEBQ or BEBQ)
A** = Self or parent-administered questionnaire (with extra validation/reliability information reported for specific target sample)
B* = Parent or self-reported questionnaire
Comparability (max 2 stars**)**
i.e. The subjects in different outcome groups are comparable, based on the study design or analysis.
1) The most important confounding factor is controlled (age, sex, education or parental adiposity).[1]
A* = Yes
B = No
2) The study controlled for any additional confounding factors.
A* = Yes
B = No
Outcome (max 3 stars**)**
1) Assessment of outcome (adiposity)
A** = Clinical assessment
B** = Record linkage
C = Self-report

D = Parent-report
2) Statistical test
A* = The statistical test used to analyse the data is clearly described and appropriate, and the measurement of the association is presented, including confidence intervals and the probability level (p value)
B = The statistical test is not appropriate, not described or incomplete.

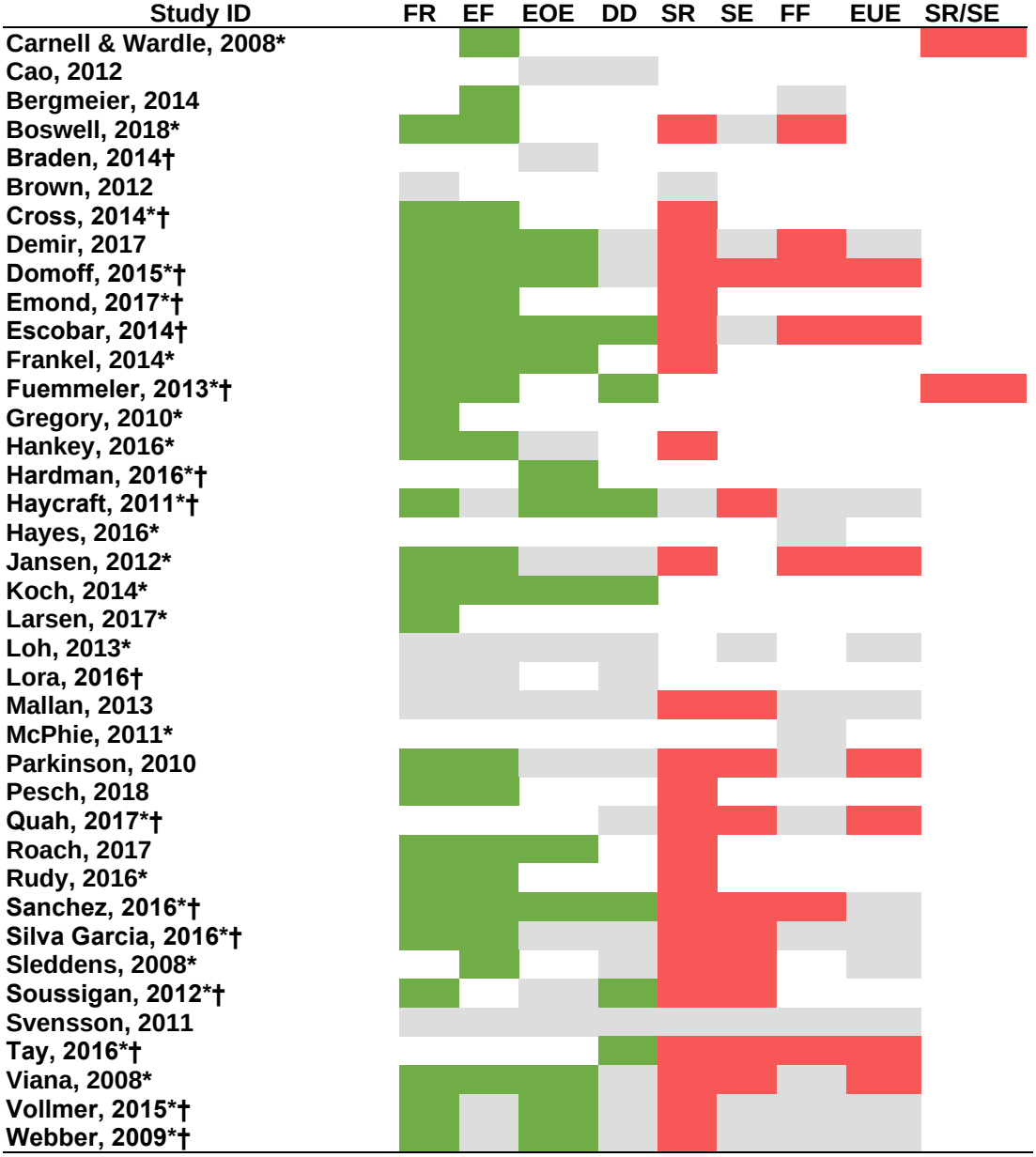
NOS modifications

This scale has been adapted from the Newcastle-Ottawa Quality Assessment Scale for cohort studies to perform a quality assessment of cohort studies for the present systematic review. We have not selected one factor that is the most important for comparability, because the variables are not the same in each study. Thus, the principal factor should be identified for each study. Where relevant, this factor could be age, sex, education or parental adiposity.

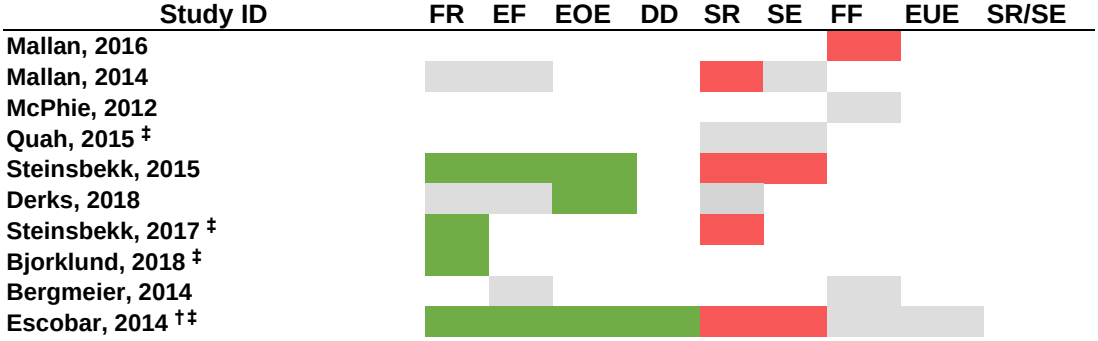
The NOS assesses three main areas of study quality, namely 1) the selection of the cohort, 2) the comparability of study analysis, and 3) the ascertainment of the outcome. Each of the three main study quality areas is rated using a star scoring system (see additional information about scoring procedure below) and the summation of the stars is used to generate the total score (maximum 10 stars). A NOS score of $\geq$7 is considered a "good" study, and this was used as a cut-off to classify good study quality (McPheeters et al. 2012).

Table S3. Results table for all CEBQ studies examining cross-sectional associations between each CEBQ scale and adiposity (n = 40), prospective associations from CEBQ scales to later adiposity (n=11), and prospective associations from adiposity to later CEBQ scales (n=5).

Cross-sectional associations of each CEBQ scale with adiposity (n = 40)



Prospective studies: association of each CEBQ scale with later adiposity (CEBQ → adiposity) (n =11)



Parkinson, 2010	

Prospective studies: association of adiposity with later CEBQ scale (adiposity → CEBQ scale) (n=5)

Study ID	FR	EF	EOE	DD	SR	SE	FF	EUE	SR/SE
Steinsbekk, 2015									
Steinsbekk, 2016									
Derks, 2018									
Steinsbekk, 2017 ‡									
van Deutekom, 2016 †									

Cross-sectional associations of each BEBQ scale with adiposity (n = 5)

Study ID	FR	EF	SR	SE	GA	SR/SE
Mallan, 2017						
Quah, 2015						
Shepard, 2015						
van Jaarsveld, 2015						
Patel, 2018						

Key: Green = positive association; Red = negative; Light grey = null; White = not measured/no data

* Indicates studies included in the meta-analysis
† Indicates studies for which authors provided additional data.
‡ When multiple time points of data are presented in the original study, the longest eligible association has been included in the table

Lipowska et al. (2018), McCartney et al. (2015) and Sanlier et al (2016) presented estimates stratified by weight status and/or gender, and therefore have not been included in this table.

Patel (2018) reported cross-sectional data exclusively
When multiple timepoints of data are presented in the original study, the longest eligible association has been included in the table above.

Table S4. Testing for linearity across weight categories (n=19)

Study ID	Test for linearity	FR	EF	EOE	DD	SR	SE	FF	EUE	SR/ SE	Weight categories used	Additional observations
Carnell & Wardle, 2008	x										Low-normal, mid-norm, high, very high	
Croker et al., 2011	x							NS	NS		UW, NW, OW, OB, Clinically OB	
de Groot et al., 2017											NW, OW	FR scores were higher for OB compared to NW (p<.001). No significant difference between NW and OB for SR, EF, DD.
dos Passos et al., 2015	x							NS	NS		NW, OW, OB, Severe OB	
Gardner et al., 2015	x		NS								NW, OB	
Ho-Urriola et al., 2014	x				NS			NS	NS		NW, OB	
Jahnke et al., 2008	x										UW, NW, OW, OB	
McCarthy et al., 2015											UW, NW, OW/OB	FR and EF were significantly higher for OW/OB children compared to UW and NW (p<.001). SR, SE and FF were significantly lower for OW/OB children compared with UW and NW (p<.001). No significant difference between weight status categories for EOE, DD and EUE.
Mosli et al. 2015											NW (<85th), OW/OB (85th>)	SR and FF were significantly lower for OW/OB children compared to NW.
Obregon et al., 2017											NW, OW, OB	FR, EOE, EF were significantly higher for OB children compared to OW and NW (p<.001). DD was significantly higher for OB children compared to NW (p<.001) but not OW. SR and SE were significantly lower for OW/OB compared to NW (p<.001).
Parkinson et al., 2010	x										BMI centile lowest, middle, highest	
Powers et al., 2006											UW, NW, at-risk for OW, OW	No significant difference between weight status categories for DD and FR.
Sanchez et al., 2016	x								NS		NW, OW, OB	
Soussignan et al., 2012											NW, OW	FR, EOE, DD were significantly higher for OB children compared to NW (p<.05). SR was significantly lower for OB children compared to NW (p<.05). No significant difference between weight status categories for EF, SE, EUE and FF.
Spence et al., 2011	x										UW, NW, at-risk for OW, OW	
Webber et al., 2009	x								NS		Thinness grade 1/2, low NW 50th centile or less, mid normal weight >50th but not OW, OW/OB	
Sandvik et al., 2018											Thinness, NW, OW, OB	Analysis showed that eating behaviours differed between the weight status groups.
Sanlier et al., 2018											UW, NW, OW, OB	FR, EOE, EF were significantly higher for OB children compared to OW, NW and UW (p<.001). SR and SE were significantly lower for OB compared to OW, NW and UW (p<.001). No significant difference between weight status categories for DD.
Boswell et al., 2018	x						NS				UW, NW, OW, OB	

Key: Green = positive association; Red = negative; Yellow = none; White = not measured/no data.