Journal of Developmental Origins of Health and Disease (2017), **8**(2), 137–148.

doi:10.1017/S2040174416000581

REVIEW

Validity of recalled *v.* recorded birth weight: a systematic review and meta-analysis

S. D. Shenkin[1,2]*, M. G. Zhang[3,4], G. Der[2,5], S. Mathur[4,6], T. H. Mina[6] and R. M. Reynolds[6]

[1] *Department of Geriatric Medicine, University of Edinburgh, Edinburgh, United Kingdom*
[2] *Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom*
[3] *Medical Student, University of Edinburgh, Edinburgh, United Kingdom*
[4] *NHS Lothian, Edinburgh, United Kingdom*
[5] *MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, United Kingdom*
[6] *University/British Heart Foundation Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom*

Low birth weight is associated with adverse health outcomes. If birth weight records are not available, studies may use recalled birth weight. It is unclear whether this is reliable. We performed a systematic review and meta-analysis of studies comparing recalled with recorded birth weights. We followed the Meta-Analyses of Observational Studies in Epidemiology (MOOSE) statement and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We searched MEDLINE, EMBASE and Cumulative Index to Nursing and Allied Health Literature (CINAHL) to May 2015. We included studies that reported recalled birth weight and recorded birth weight. We excluded studies investigating a clinical population. Two reviewers independently reviewed citations, extracted data, assessed risk of bias. Data were pooled in a random effects meta-analysis for correlation and mean difference. In total, 40 studies were eligible for qualitative synthesis ($n = 78,997$ births from 78,196 parents). Agreement between recalled and recorded birth weight was high: pooled estimate of correlation in 23 samples from 19 studies ($n = 7406$) was 0.90 [95% confidence interval (CI) 0.87–0.93]. The difference between recalled and recorded birth weight in 29 samples from 26 studies ($n = 29,293$) was small [range −86–129 g; random effects estimate 1.4 g (95% CI −4.0–6.9 g)]. Studies were heterogeneous, with no evidence for an effect of time since birth, person reporting, recall bias, or birth order. In *post-hoc* subgroup analysis, recall was higher than recorded birth weight by 80 g (95% CI 57–103 g) in low and middle income countries. In conclusion, there is high agreement between recalled and recorded birth weight. If birth weight is recalled, it is suitable for use in epidemiological studies, at least in high income countries.

Received 27 May 2016; Revised 26 September 2016; Accepted 29 September 2016; First published online 25 October 2016

Key words: birth weight, meta-analysis, systematic review, validation

Introduction

Birth weight is an important marker of current and future health, and has been used in many epidemiological studies of determinants of health and disease from childhood through adulthood to old age.[1,2] Some studies have recorded birth weight directly in official records,[3] but many studies rely on recalled birth weight reported by the participants or their mothers.[4] Several studies have found that maternal recall is fairly accurate, even years after the birth,[5,6] but to our knowledge there has been no systematic review to establish whether this finding is consistent across all published studies. This systematic review and meta-analysis of published observational studies aimed to determine the agreement between birth weight recalled by parent or self any time after birth, and the actual birth weight recorded in official records.

*Address for correspondence: S. D. Shenkin, Department of Geriatric Medicine, University of Edinburgh, 51 Little France Crescent, EH16 4SB, Edinburgh, United Kingdom.
(Email susan.shenkin@ed.ac.uk)

Methods

Data sources

We followed the Meta-Analyses of Observational Studies in Epidemiology (MOOSE) guidelines for the conduct,[7] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for the reporting,[8] of this systematic review. M.G.Z. performed the literature search on MEDLINE, EMBASE and Cumulative Index to Nursing and Allied Health Literature (CINAHL) from inception to May 2015 using terms as both keywords and indexing (MeSH) terms: birth weight AND (mental recall OR self-report) AND (recorded OR actual OR verified) (full strategy: Supplementary material 1). We also searched reference lists and performed a forward citation search of all included papers.

Study selection

We included studies in the systematic review which addressed the question: 'Does recalled birth weight correlate with recorded birth weight?' We included both self and parental recall, with no restriction on time from birth. We excluded studies that did not report a 'gold standard' for birth weight (recorded in official document, e.g. birth certificate or birth register). We excluded individuals with specific mental or physical illnesses to ensure results were applicable to the general population, but included control groups if these were reported separately. We excluded studies that selected participants on the basis of abnormal births (e.g. low birth weight or preterm) as a high-risk pregnancy or birth may affect frequency of measurement, and influence maternal recall, but included studies that included all unselected births. We excluded studies which only categorized birth weight into two or three categories. There were no exclusions by age, sex, socioeconomic status, ethnicity or country, or language of publication. M.G.Z., S.M. and T.H.M. independently identified studies for inclusion, resolving any disagreements by consensus, and/or discussion with S.D.S. and R.M.R. The protocol is available by contacting the authors.

Data extraction

M.G.Z. and T.H.M. independently extracted relevant information on study characteristics (Table 1), and results (Table 2) directly to Excel spreadsheets. This included factors which may influence recall of birth weight, that is time since birth, method of recall (questionnaire or interview) and parity. Each paper was assessed qualitatively for major sources of bias or confounding.

Where data were not published, we contacted authors twice by email and post. We received a response with data from two (Sou,[9] data included; Tehranifar,[10] some required data not collected, therefore not included) and two further stating that data were not available. If there was no response, we estimated values using data or figures in the paper, for example, the standard deviation of mean differences.[11–13] Where studies reported some form of correlation between the measures (Pearson's r, Spearman's ρ, ICC or κ) this was used in the main analysis if calculated on continuous (individual) birth weight measures, but not if calculated using categories of birth weight. Where more than one measure was reported, we used Pearson's r.

Where no correlation measure was reported, we used the summary estimate from the other studies as described below. Jaspers *et al.*[14] reported an upper CI which appeared too large (0.16 pounds = 80 g), given the mean difference of 25 g and the lower interval of 10 g. We contacted the author but have not received a reply, so have used an upper CI of 40 g.

The main quality assessment was the risk of bias in recall of birth weight due to access to the gold standard (e.g. birth certificate). We categorized risk of bias as high if the subjects had access to this document at the time of the study, low if they did not have access, or if this was unclear (i.e. not reported, but possible, for example, telephone interview where parent would have had access to birth records kept at home).

Meta-Analysis

The meta-analysis was conducted with Comprehensive Meta Analysis V3.3 (Biostat, Englewood, CO, USA) using inverse variance weighting and the method of moments for random effects.[15] This means that the impact of the sample size is proportional to its square root. The main analysis summarized the mean difference in grams between measured and recalled birth weight.

To accurately calculate the variance of the difference requires knowledge of the correlation between recalled and measured birth weight.

The first step was to produce a summary estimate of the correlation from those studies that reported it. The summary estimate was then used in the main analysis for those studies that did not report a correlation.

A preliminary fixed effects analysis revealed high levels of heterogeneity ($I^2 = 80\%$); we therefore report summary effects from random effects models.

Sensitivity analyses were conducted for (1) recall bias (only including studies without recall bias); (2) time elapsed since birth (only including those >1 year); (3) parity correction (only including studies which corrected for parity); (4) studies using estimated values; (5) study sample size (omitting the two largest studies and conducting a leave-one-out analysis); (6) the estimated correlation between measures (using the values of the 95% CI in place of the summary estimate).

Subgroup analyses were conducted for (1) self *v.* parental recall; (2) metric *v.* imperial units of measurement; (3) high *v.* low and middle income countries. The first two were pre-specified, whereas the third was *post-hoc*, suggested by a reviewer. Meta-regression was used to explore further significant subgroup differences.

Results

From 962 abstracts, 147 full-text articles were assessed (Fig. 1), and 40 studies were included in the qualitative synthesis (Table 1),[6,9–14,16–48] with 23 samples from 19 studies included in the meta-analysis of correlation, and 29 samples from 26 studies included in the meta-analysis of mean difference.[6,9,11–14,16,17,19,20,25,26,28–32,34–45,47] Only four non-English papers were identified, and from non-expert translation three did not appear eligible, and one[23] was included in narrative review only (Table 1).

Qualitative synthesis

In total, 40 studies were eligible for inclusion in the systematic review (Table 1). They were heterogeneous: size in the recalled group ranging from 14 to 46,637 (median 257), the year of publications ranging from 1935 to 2013; the majority from the United States (18 studies) and Europe (13 studies); birth information was mostly reported by mothers (31 samples), self (eight samples) or either parent (five samples). Two studies reported both mother and self-report.[30,41] The time to recall for parental report varied from 3 weeks to 96 years, and for self-report from 27 to 78 years. Data collection was by interview (20 studies, including three by telephone), questionnaire (17 studies) or both. Recorded data were from clinical (hospital or birth register) records (33 studies), birth certificates (four studies), or research databases collected at birth (four studies).

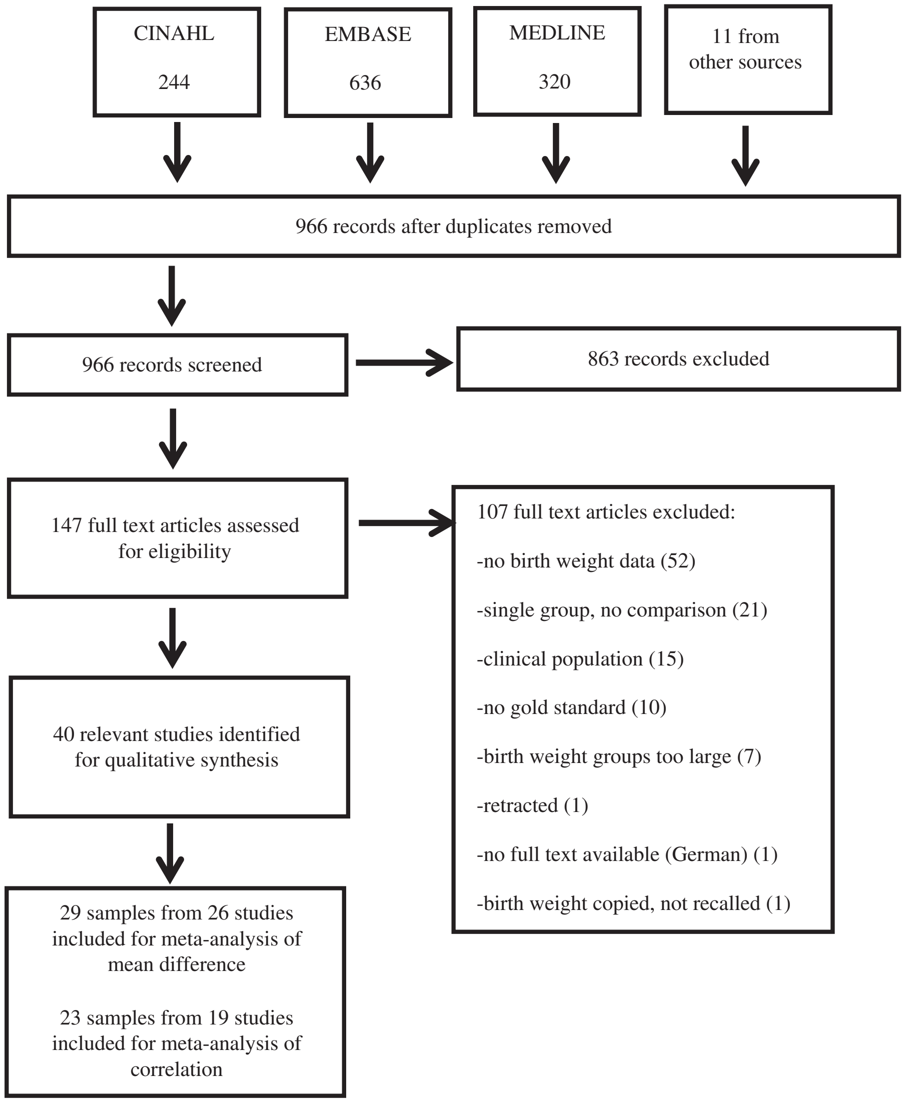

Fig. 1. Flow [Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)] diagram of included studies. CINAHL, Cumulative Index to Nursing and Allied Health Literature

The majority reported metric measures (g); where imperial measures were used we converted to metric (1 oz = 28 g). Note one study used 'Dutch modern pounds' = 500 g.[14]

There were 10 samples from nine studies,[10,20,23–27,46,48] which did not provide data for meta-analysis. These included from 47 to 2552 mothers (median 99) (Table 1) and generally reported good agreement within birth weight categories (Table 2), with over 50% of participants reporting agreement within 25 g (1 oz) (20,23), and 70–90% agreeing within 100 g.[20–24,27,47,48] The majority of studies were small ($n < 200$), with an unclear risk of bias (i.e. most studies did not report whether or not the informant had access to a recorded birth weight). Bat-erdene *et al.*[48] ($n = 2552$) estimated maternal recall at up to 3 months compared with electronic health records and found that 11.1% had exact recall, and 88.4% within 50 g; Victora *et al.*[23] ($n = 1800$) in Brazil at 9–15 months found 60% of mothers recalled the exact weight.

The largest study by far was eligible for meta-analysis: Gayle *et.al.*[47] ($n = 46{,}637$), followed up participants in the Tennessee Women, Infants and Children Supplemental Feeding Program in the United States, and found 70.6% mothers had exact recall, and 89% within 28 g. This study included 20% preterm, and 7.4% low birth weight, but we did not exclude this study as these groups were not intentionally oversampled. The time to recall was not reported, though they reported that there was no difference in recall if child's age was greater or <1 year. There was no access to the electronic health record. Lower accuracy was associated with infant's low birth weight, poor birth outcome, poorer education, black race, single marital status and age <18 years. Mothers reported a 0.2 oz (6 g) lower mean birth weight compared with birth certificates.

Most studies do not report the proportion who were unable to recall birth weight: in Allen *et al.*[40] this was 47% (Table 2). In summary, included studies find that almost 90% of mothers recall birth weight to within 1–2 oz (Table 2).

Table 1. *Descriptive data of studies included in systematic review of recalled v. recorded birth weight (in order of publication)*

Year	First author	Country	Study population (*n*)	Recall time (years)	Data collection	Birth records	Parity correction	Unit used	Recall bias	Additional notes
1935	Pyles[a]	USA	223 mothers	21 months	I	Nurse records	No	Oz	Low	–
1967	Donoghue[a]	UK	69 mothers	43 months	I	Hospital records	No	Oz	Low	–
1968	Bailit	USA	372 births from 136 mothers, 36.99 (6.55) years	10.10 (4.03)	Q	Hospital records	Yes	lb	Low	–
1972	Porteous[a]	USA	298 mothers	0–3 (29), 4–7 (94), 8–10 (107), 11–14 (68)	I	Birth certificate	No	Oz	Low	–
1976	Hoekelman[b]	USA	59 mothers, 16–38 years old	9 months	I	Clinical records	Yes	Oz	Low	–
1984	Axelsson	Sweden	511 mothers	0–7	Q	Medical Birth Register, Sweden	No	g	Unclear	–
1984	Oates	Kenya	47 mothers	23–42	I	Hospital records	No	g	Unclear	–
1985	Victora	Brazil	1800 mothers	9–15 months	Q	Hospital records	No	g	Unclear	In Portuguese
1986	Eaton-Evans	Australia	81 mothers	1–10	I	Clinical records	Yes	Oz/lb	Low	–
1987	Burns[a]	USA	127 mothers, 42.9 (6.4) years old	16.1 (2.4)	I	Hospital delivery record	Yes	lb	Low	–
1987	Seidman[a]	Israel	97 mothers of 662 children, 27–68 (mean 35) years old	4–6, 10–14, 15–19, 20–23	I	Hospital records	Yes	g	Low	All mothers had >7 children
1988	Gayle[a]	USA	46,637 mothers	NR	Q	Birth certificate	Yes	Oz	Low	Data recorded from Supplemental Feeding Program; includes 20% preterm and 7.4% LBW
1991	Wilcox	USA	104 mothers	≤2 (56%), >2 (48%)	I	Medical records	Yes	g	Low	–
1993	Diaz[a]	Peru	128 mothers	1–3 months	I	Delivery records in hospital	No	g	Low	–
1994	Lumey[a]	Holland	861 mothers, 43.3 (1.1) years old	1–29	I	Hospital birth records	Yes	g	Low	–
1995	Pless[a]	Canada	288 parents (91% mothers), 288 children	1–13	Q	Medical records	Yes	g	Low	–
1996	Troy[b]	USA	220 mother-self pairs	Self: 27–44 years	Q	State Birth Records	No	lb	Unclear	
	Troy[b]		As above	Mother: 46 to 96 years		As above				–
1997	Olson[a]	USA and Canada	558 (matched controls for leukemia <18 months)	Up to 8 years	I (T)	Medical records	No	g	Low	
1997	Gaskin[a]	Jamaica	243 mothers, 2 child age groups	3–4	NR	Hospital records	No	Oz	Unclear	
	Gaskin[a]		As above	7–8		As above				–
1998	Sanderson	USA	161 controls for breast cancer: self	<45	I (T)/Q	Birth certificate	No	g	Low	
	Sanderson	USA	106 mothers of controls: child	NR	I (T)/Q	Birth certificate	No	g	Low	
1998	Lederman[a]	USA	144 mothers, 26 (4.8) years old	3 weeks	I	Clinical records	No	g	Low	–
1999	Tomeo[a]	USA	154 mothers, average age 57	32	Q	National collaborative perinatal project (NCPP) records	No	g	Unclear	–
2000	Gofin[a]	Israel	259 mothers, 19 years old and above	6	I	Medical records	Yes	g	Unclear	–
2000	Kemp[a]	UK	73 persons, self-report	64 (3.2)	Q	Birth and clinical records	No	kg	Unclear	–
2000	Andersson[a]	Sweden	192 persons, self-report	44 (92), 52 (58), 56 (35), 60 (7)	Q	Delivery records	Yes	kg	Unclear	–
2000	O'Sullivan[a]	UK	649 parents (not distinguished)	6–15	Q	Birth records	No	g	Unclear	–
2000	Walton[a]	UK	873 parents (not distinguished)	12 and 15	Q	Birth records on the Child Health System	Yes	Oz/lb	Unclear	–
2002	Allen[b]	UK	128 persons, self-report	43–50 ($n = 81$) and 63–78 ($n = 40$)	Q	Midwife's register	Yes	Oz	Unclear	–

2005	Tate[a]	UK	11,890 mothers	9 months	I	Birth Registry and Millennium Cohort Study	Yes	lb/oz/kg	Low	–
2006	Catov[a]	USA	40 mothers, mean age 80 years old: 14 first births	Average 57 (5)	I	Birth records	Yes	Unclear	High	
	Catov[a]		As above: 26 subsequent births	Average 57 (5)		As above				–
2006	Lucia[a]	USA	644 mothers-self pairs, recall I by mother and adolescent	17 (mothers)	I	Hospital record	Yes	g	Low	
	Lucia[a]		As above	17 (self)		As above				–
2006	Sou[a]	Taiwan	107 mothers, mean age 37.5 (4.7) term deliveries	3–9	I	Medical records	Yes	g	Low	
2007	Araujo[a]	Brazil	3426 mothers	11	I	Data from the 1993 Pelotas Birth Cohort Study measured at birth	Yes	g	Low	Norway
2008	Adegboye[a]	Denmark	1271 mothers, 40.4 (5.3) years old	8–11 (68% $n = 971$); 14–18 (32%, $n = 456$)	Q	Danish Medical Birth Register (DBR)	Yes	g	Low	–
2009	Tehranifar	USA	223 persons, self-report	38–46	Q	Birth records	No	lb	Unclear	–
2010	Wodskou[a]	Denmark	441 nurses, mean age 53.5 years, self-report (of 517 with recall data and 925 with recorded data, subgroup of Danish Nurses Cohort Study)	44–69	Q	Birth record part of Copenhagen Schools Health Records Register	No	g	Unclear	–
2010	Jaspers[a]	Holland	1879 parents	11–12	Q, I	Medical records	No	Modern lb' (500 g)	Low	Collected in units of lb (500 g)
2012	Boeke[a]	Colombia	279 mothers, 20 years old and above	5–12 (mean 8.6, S.D. 1.6)	Q	Birth records in hospitals in Bogota	Yes	g	Unclear	–
2012	Lule[a]	Uganda	265 mothers, hospital births	4–7	I	EMaBS (Entebbee mother and baby study) data recorded at birth	Yes	g	Low	–
2013	Bat-erdene	Canada	2552 mothers	0.33	Q	Electronic Health Records (Calgary)	Yes	kg	Unclear	–

BW, birth weight; ICC, intraclass correlation; LBW, low birth weight; NR, not recorded.
Where data reported in oz, converted to g (1 oz = 28 g).
Where mean difference is negative, recorded BW larger than recalled BW.
[a]Included in meta-analysis of mean difference.
[b]Included in meta-analysis of correlation only.

Table 2. *Results of studies included in systematic review of recalled v. recorded birth weight (in order of publication)*

Year	First author	Recalled mean birth weight (g)	Recalled S.D.	Recalled sample size	Recorded mean birth weight (g)	Recorded S.D.	Recorded sample size	Correlation	Categorical findings	Author's remark on the quality of birth weight recall (systematic review only)
1935	Pyles[a]	3466.01	594.5	223	3476.2	578.6	252	0.96	MD −10.2 g, 59% within 50 g	
1967	Donoghue[a]	3260.2	538.6	69	3260.2	481.9	69	0.96		
1968	Bailit	NR	NR	372	3379	549	372	NR	51% of birth weight were incorrect by >28 g. 5 of these, 58% were under-estimates. MD = 104 g	Varies according to parity, implication unclear (for ⩾4 children the error was up to 259 g)
1972	Porteous[a]	NR	NR	298	NR	NR	298	NR	MD = − 12.3 g (S.D. 97.8 g)	
1976	Hoekelman[b]	3209.2	NR	59	3265.9	NR	59	0.86	68% within 28 g	32% inaccurate (recall-gold standard difference >100 g)
1984	Axelsson	NR	NR	551	NR	NR	551	NR	72% of recalled birth weight were different than recorded ones. 7% had differences of >100 g. 33% of errors were due to rounding to the nearest hundreds of grams	28% inaccurate (recall-gold standard difference >100 g)
1984	Oates	NR	NR	24	NR	NR	24	NR	50% of recalls were within 30 g, 71% within 100 g, 29% had discrepancies of >200 g	50% inaccurate (recall-gold standard difference >100 g)
1985	Victora	NR	NR	1800	NR	NR	1458	NR	60% recalled the exact weight, 80% recalled within 100 g, 90% within 250 g	80% inaccurate (recall-gold standard difference >100 g)
1986	Eaton-Evans	NR	NR	81	NR	NR	81	NR	75% of recalls were within 100 g. 4% had discrepancies of >200 g	27% inaccurate (recall-gold standard difference >100 g)
1987	Burns[a]	3447.3	589.7	127	3492.7	589.7	127	0.94		
1987	Seidman[a]	NR	NR	662	NR	NR	662	ICC = 0.71	MD = 95 g (S.D. 185); 41% within 10 g, 75% within 100 g, 87% within 200 g, 92% within 300 g	
1988	Gayle[a]	NR	NR	46637	NR	NR	72245	NR	Maternal report 6 g lower; 70.6% exact, 89% within 28 g, 90.6% within 56 g, 91.6% within 84 g, 82.5% within 112 g, 95.5% within 226 g	11% inaccurate (recall-gold standard difference >1 ounce)
1991	Wilcox	NR	NR	125	NR	NR	104	NR	68% within 100 g, 79% within 200 g	31.7% inaccurate (recall-gold standard difference >100 g)
1993	Diaz[a]	3390	NR	128	3400	NR	128	NR		
1994	Lumey[a]	3338	666	1297	3342	586	1297	NR	Bland Altman scatterplot shown	
1995	Pless[a]	3441	562	271	3438	565	271	NR	Cohens $k = 0.81$ (categorical). 73.4% within 50 g, 84% within 150 g	
1996	Troy[b]	NR	NR	220	NR	NR	220	0.74	Self	~30% inaccurate (also see correlation)
	Troy[b]	NR	NR	220	NR	NR	220	0.85	Mother	
1997	Olson[a]	NR	NR	558	NR	NR	558	0.978	95% CI 0.974–0.982, κ 0.91, MD −10.6 g (95% CI −10.29 to −0.83: includes both cases and controls) ('no meaningful difference if stratified by case/control status'). Also reports birth weight reported as continuous, and <3000 g, 3000–3499 g, 3500–2999 g, ⩾4000 g	
1997	Gaskin[a]	3030	590	111	2980	540	111	0.85	Age 3–4	
	Gaskin[a]	3350	520	132	3280	470	132	0.77	Age 7–8	
1998	Sanderson	NR	NR	161	NR	NR	161	$\rho = 0.80$[c]	Spearman correlation between category, categories ⩽2500, 2500–2999, 3000–3499, 3500–3999, ⩾4000 g); ~20% underestimated BW by one category	
	Sanderson	NR	NR	106	NR	NR	106	$\rho = 0.84$[c]		
1998	Lederman[a]	3447	447	144	3452	450	144	NR		
1999	Tomeo[a]	2852	565	154	2877	508	154	0.9		

2000	Gofin[a]	3194.6	506.6	319	3206.4	510	259	NR	κ = 0.71 using 500 g categories, 58% within 100 g, 80% within 500 g	
2000	Kemp[a]	3360	860	73	3420	580	73	0.64	κ = 0.43 if data categorized as <2.5 kg, 2.5–3.5 kg, >3.5 kg	
2000	Andersson[a]	NR	NR	192	NR	NR	192	0.76		
2000	O'Sullivan[a]	NR	NR	649	3380	NR	649	0.95	40% within 10 g, 76% within 50 g, 85% within 100 g, 90% within 200 g	
2000	Walton[a]	3375	542.3	873	3378	526.8	873	0.9		
2002	Allen[b]	NR	NR	244	NR	NR	244	0.86	25% of recall are within 4 ounces, 28% reported inaccurately	25% within 113.4 g, 28% inaccurate, 47% responded 'don't know'
2005	Tate[a]	3360	580	11,890	3361	570	11,890	NR	82% within 30 g, 92% within 100 g	
2006	Catov[a]	NR	NR	14	NR	NR	14	ICC = 0.96	first birth (not stated how ICC was calculated 'between recalled and documented')	
	Catov[a]	NR	NR	26	NR	NR	26	ICC = 0.59	Subsequent birth	
2006	Lucia[a]	NR	NR	564	NR	NR	644	0.97	54% within 15 g, 87.1% within 250 g, Bland Altman plot shown	
	Lucia[a]	NR	NR	486	NR	NR	644	0.83	24.1% within 15 g, 61.3% within 250 g	
2006	Sou[a]	NR	NR	107	3311.8	351.6	107	0.89	Average discrepancy of 30.6 g over reported, mean error = 72.5 g, s.d. 125.7 g	
2007	Araujo[a]	3197	574	3426	3177	524	3426	NR	κ for < or > 2500 g = 0.73; Bland Altman plot shown; 32.6% identical, 75.1% within 200 g	
2008	Adegboye[a]	NR	NR	1271	3388	567.1	1271	0.97	Overall ICC = 09.4, Bland Altman plots shown, 96.4% within 285 g; MD = − 0.02 (s.d. 142.4)	
2009	Tehranifar	NR	NR	223	3139.2	NR	223	0.67	73% correctly estimated birth weight category	Moderate to good recall (sensitivity ~ 73%, weighted κ = 0.67)
2010	Wodskou[a]	3249.1	671.1	517	3321.8	567.6	925	0.83	MD = 72.7 g between all records recalled and recorded; n = 441 for recalled and recorded: Bland Altman plot shown, MD = 20.9 g; 40.4% within 50 g, 54.4% within 100 g, 73.9% within 250 g; 98% within 1000 g	
2010	Jaspers[a]	3400	600	1691	3450	550	1691	NR	Collected to nearest 'modern lb' (500 g) MD = 25 g (recorded larger); 95% within 600 g; Bland Altman plot given	
2012	Boeke[a]	3106	739	279	2977	462	279	NR		
2012	Lule[a]	3280	680	303	3210	500	265	ICC = 0.64[c]	ICC calculated on <2.5; 2.5–4, >4.0 kg; 14% exact, 34% within 100 g	
2013	Bat-erdene	NR	NR	2552	NR	NR	2552	NR	11.1% difference 0 g; 88.4% within 50 g; 91.7% within 100 g, 93.7% within 150 g; 94.5% within 200 g	Excellent (11.6% exact recall, 91.7% recalled within 200 g, specificity >85%)

MD, mean difference.
Where data reported in oz, converted to g (1 oz = 28 g).
Where MD is negative, recorded BW larger than recalled BW.
[a]Included in meta-analysis of mean difference.
[b]Included in meta-analysis of correlation only.
[c]Correlation coefficient/κ reported for categorical analysis therefore not included in meta-analysis.

Meta-analysis

We included 23 samples from 19 studies (total $n = 7406$) in the meta-analysis of correlation, and 29 samples from 26 studies (total $n = 72{,}114$) in the meta-analysis of differences in birth weight (Table 1 and 2): three studies[13,32,41] had two sets of data which allowed separate analysis: two age groups;[32] first *v.* subsequent births;[13] maternal *v.* self recall[41] (Table 2). Sample size ranged from 14 to 46,637, median 265.

Correlation

There was a strong correlation between recalled and recorded birth weight, estimated as 0.90 (CI 0.86–0.93) (Fig. 2). This estimate of the correlation was used in the main analysis for studies that did not report a correlation.

Differences in absolute birth weight

The absolute effect size of the difference in birth weight between recalled and recorded was very small, not statistically significant, and unlikely to be clinically important: 1.4 g (−4.0 to 6.9 g) (Fig. 3).

Sensitivity analysis

Sensitivity analyses to assess the effect of – (1) recall bias; (2) time elapsed since birth; (3) parity correction; (4) studies using estimated values – all showed little effect on the results (Supplementary Figs 1–4). Leaving out the two very large studies – Gayle ($n = 46{,}637$) and Tate ($n = 11{,}890$) – yielded a summary estimate of 5.82 g (−4.36, 16.00). A leaving one out analysis showed that no other study affected the summary estimate by more than 2 g (Supplementary Fig. 5). For eight studies, we used a summary estimate of the correlation. We therefore also performed sensitivity analyses in which we substituted the upper and lower 95% limits of the estimated correlation (0.93 and 0.85) for those studies that did not report one. The results (mean difference, 95% CI) are: 1.88 g (−3.64, 7.41) and 0.96 g (−4.50, 6.39), for the upper and lower limit, respectively.

Subgroup analyses

Subgroup analysis by informant and units of measurement yielded subgroup estimates that were not significantly different (Supplementary Figs 6 and 7). In contrast, the analysis by

Study name	Statistics for each study				
	Correlation	Lower limit	Upper limit	Z-Value	p-Value
1935_Pyles	0.960	0.949	0.969	30.706	0.000
1967_Doneghue	0.960	0.936	0.975	15.809	0.000
1976_Hoekelm	0.860	0.774	0.915	9.679	0.000
1987_Burns	0.940	0.916	0.957	19.354	0.000
1996_Troy1	0.740	0.674	0.795	14.001	0.000
1996_Troy2	0.850	0.800	0.888	15.538	0.000
1997_Gaskin1	0.850	0.789	0.895	13.054	0.000
1997_Gaskin2	0.770	0.690	0.831	11.589	0.000
1997_Olson	0.978	0.974	0.981	52.992	0.000
1999_Tomeo	0.910	0.878	0.934	18.771	0.000
2000_Andersson	0.760	0.693	0.814	13.696	0.000
2000_Gofin	0.890	0.862	0.913	22.751	0.000
2000_Kemp	0.640	0.481	0.758	6.343	0.000
2000_O'Sullivan	0.950	0.942	0.957	46.558	0.000
2000_Walton	0.900	0.887	0.912	43.424	0.000
2002_Allen	0.860	0.807	0.899	14.518	0.000
2006_Catov1	0.970	0.905	0.991	6.939	0.000
2006_Catov2	0.590	0.263	0.796	3.250	0.001
2006_Lucia1	0.970	0.965	0.975	49.557	0.000
2006_Lucia2	0.830	0.800	0.856	26.112	0.000
2006_Sou	0.890	0.843	0.924	14.501	0.000
2008_Adegboye	0.970	0.967	0.973	74.505	0.000
2010_Wodskou	0.830	0.809	0.849	36.077	0.000
	0.899	0.856	0.929	15.466	0.000

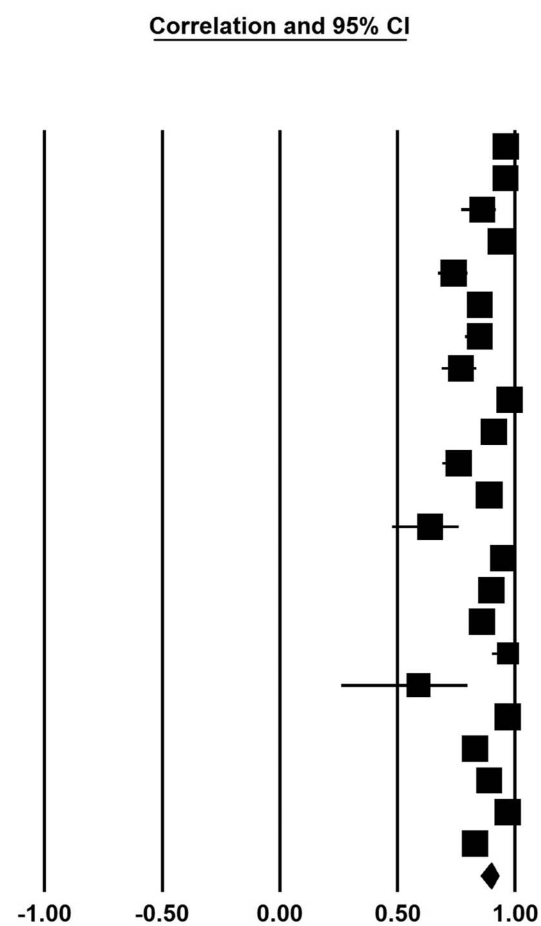

Fig. 2. Meta-analysis of correlations.

https://doi.org/10.1017/S2040174416000581 Published online by Cambridge University Press

Study name	Statistics for each study					
	Difference in means	Lower limit	Upper limit	Z-Value	p-Value	Total
1935_Pyles	10.2	-11.7	32.1	0.915	0.360	223
1967_Doneghue	0.0	-36.5	36.5	0.000	1.000	69
1972_Porteous	-12.3	-23.4	-1.2	-2.172	0.030	298
1987_Burns	-36.9	-72.9	-0.8	-2.006	0.045	127
1988_Gayle	-6.0	-6.9	-5.1	-12.957	0.000	46637
1993_Diaz	-10.0	-505.2	485.2	-0.040	0.968	128
1994_Lumey	4.0	-11.9	19.9	0.493	0.622	1297
1995_Pless	3.0	-52.2	58.2	0.106	0.915	288
1997_Gaskin1	50.0	-7.7	107.7	1.699	0.089	111
1997_Gaskin2	70.0	12.0	128.0	2.365	0.018	132
1998_Lederman	-4.2	-17.6	9.2	-0.615	0.539	144
1999_Tomeo	-25.0	-64.5	14.5	-1.241	0.215	154
2000_Andersson	4.7	-48.1	57.5	0.175	0.861	192
2000_Gofin	11.8	-16.0	39.6	0.831	0.406	259
2000_Kemp	60.0	-91.7	211.7	0.775	0.438	73
2000_O'Sullivan	-11.0	-23.5	1.5	-1.730	0.084	649
2000_Walton	-3.0	-18.9	12.9	-0.370	0.711	873
2005_Tate	-1.0	-3.7	1.7	-0.722	0.470	11890
2006_Catov1	-44.0	-115.7	27.7	-1.203	0.229	14
2006_Catov2	-86.0	-218.3	46.3	-1.274	0.203	26
2006_Lucia1	-2.9	-20.3	14.5	-0.327	0.743	564
2006_Lucia2	-24.5	-73.6	24.6	-0.979	0.328	486
2006_Sou	72.5	48.7	96.3	5.966	0.000	107
2007_Araujo	20.0	4.3	35.7	2.501	0.012	3426
2008_Adegboye	-0.2	-8.0	7.6	-0.050	0.960	1271
2010_Jaspers	-25.0	-40.0	-10.0	-3.269	0.001	1691
2010_Wodskou	-20.9	-59.7	17.9	-1.057	0.290	441
2012_Boeke	129.0	84.2	173.8	5.645	0.000	279
2012_Lule	70.0	6.6	133.4	2.164	0.030	265
	1.4	-4.0	6.9	0.515	0.606	

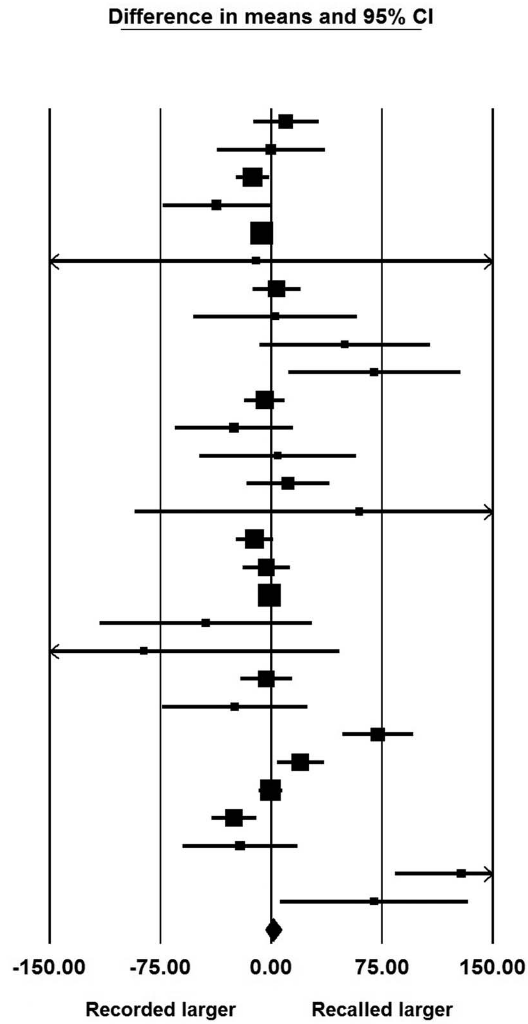

Fig. 3. Difference between recalled and recorded birth weight (g).

country income category revealed a striking difference. Low and middle income countries appear to overestimate birth weight by around 80 g (57,103) (Fig. 4). The income categorization explained 77% of between study variance, but unexplained variance was still moderately high ($I^2 = 48\%$).

Risk of bias

Most studies were observational cohort studies of good quality with little evidence of major source of biases or confounding factors. Some studies analyzed subgroups to determine if there were subgroups with higher or lower errors. Inclusion and exclusion criteria were generally not well reported. The main source of bias was the possibility that participants were not blinded to the recorded birth weight (e.g. birth certificate), and for most studies it was unclear whether or not participants had access to such records. One excluded study[49] explicitly asked parents to copy results from a personal child health record. Results were essentially unchanged if we excluded studies where access to the birth weight record was possible (difference in means −0.04 g (CI −5.6–5.5 g).

Discussion

This systematic review of 40 studies (total $n = 78{,}997$ births) and meta-analysis in 29 samples from 26 studies (total $n = 72{,}114$) shows that recalled birth weight has excellent agreement with recorded birth weight: pooled estimate of correlation in 23 samples from 19 studies (total $n = 7406$ births) was 0.90 (95% CI 0.86–0.93), with a small absolute difference: range from −86 to +129 g; random effects estimate 1.4 g (95% CI −4.0–6.9 g). There was no evidence for an effect of self or parental recall, age at recall or time elapsed since birth event on the validity of recalled birth weight. There was, however, evidence of higher recalled birth weight of 80 g (95% CI 57–103 g) in low or middle income countries, in *post-hoc* analysis.

The majority of the studies included reported high agreement, with a small (clinically insignificant) absolute difference.

Subgroup analysis: High vs Low & Middle Income Countries

Group by Income	Study name	Difference in means	Lower limit	Upper limit	Z-Value	p-Value	Total
HIC	1935_Pyles	10.2	-11.7	32.1	0.915	0.360	223
HIC	1967_Doneghue	0.0	-36.5	36.5	0.000	1.000	69
HIC	1972_Porteous	-12.3	-23.4	-1.2	-2.172	0.030	298
HIC	1987_Burns	-36.9	-72.9	-0.8	-2.006	0.045	127
HIC	1988_Gayle	-6.0	-6.9	-5.1	-12.957	0.000	46637
HIC	1994_Lumey	4.0	-11.9	19.9	0.493	0.622	1297
HIC	1995_Pless	3.0	-52.2	58.2	0.106	0.915	288
HIC	1998_Lederman	-4.2	-17.6	9.2	-0.615	0.539	144
HIC	1999_Tomeo	-25.0	-64.5	14.5	-1.241	0.215	154
HIC	2000_Andersson	4.7	-48.1	57.5	0.175	0.861	192
HIC	2000_Gofin	11.8	-16.0	39.6	0.831	0.406	259
HIC	2000_Kemp	60.0	-91.7	211.7	0.775	0.438	73
HIC	2000_O'Sullivan	-11.0	-23.5	1.5	-1.730	0.084	649
HIC	2000_Walton	-3.0	-18.9	12.9	-0.370	0.711	873
HIC	2005_Tate	-1.0	-3.7	1.7	-0.722	0.470	11890
HIC	2006_Catov1	-44.0	-115.7	27.7	-1.203	0.229	14
HIC	2006_Catov2	-86.0	-218.3	46.3	-1.274	0.203	26
HIC	2006_Lucia1	-2.9	-20.3	14.5	-0.327	0.743	564
HIC	2006_Lucia2	-24.5	-73.6	24.6	-0.979	0.328	486
HIC	2007_Araujo	20.0	4.3	35.7	2.501	0.012	3426
HIC	2008_Adegboye	-0.2	-8.0	7.6	-0.050	0.960	1271
HIC	2010_Jaspers	-25.0	-40.0	-10.0	-3.269	0.001	1691
HIC	2010_Wodskou	-20.9	-59.7	17.9	-1.057	0.290	441
HIC		-4.2	-7.9	-0.5	-2.199	0.028	
LMIC	1993_Diaz	-10.0	-505.2	485.2	-0.040	0.968	128
LMIC	1997_Gaskin1	50.0	-7.7	107.7	1.699	0.089	111
LMIC	1997_Gaskin2	70.0	12.0	128.0	2.365	0.018	132
LMIC	2006_Sou	72.5	48.7	96.3	5.966	0.000	107
LMIC	2012_Boeke	129.0	84.2	173.8	5.645	0.000	279
LMIC	2012_Lule	70.0	6.6	133.4	2.164	0.030	265
LMIC		79.9	56.8	102.9	6.796	0.000	
Overall		-2.0	-5.7	1.6	-1.089	0.276	

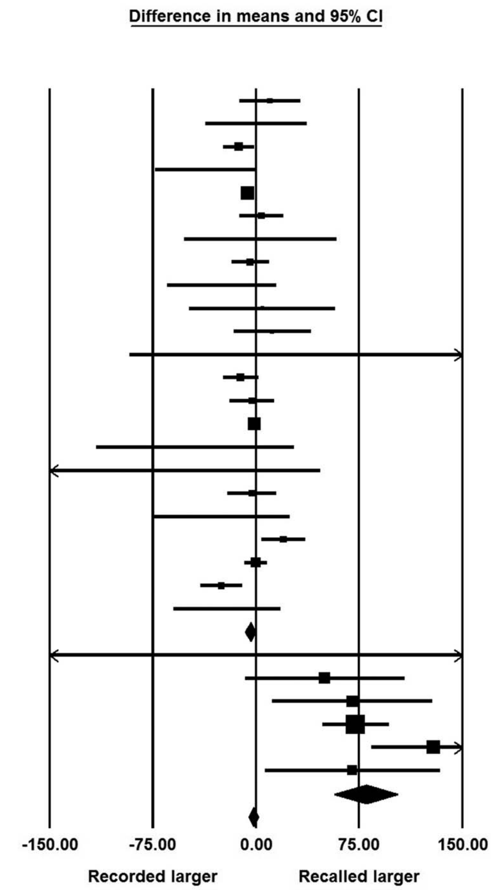

Fig. 4. Subgroup analysis of high *v.* low and middle income countries. HIC, high income country; LMIC, Low/middle income country.

In studies which reported findings in categories, rather than absolute values, over 50% of participants reported agreement within 25 g (1 oz). If a 100 g error was tolerated, most studies reported agreement between 70 and 90%. Some of the differences may be due to reporting (rounding) errors: if reporting in imperial measures to the nearest ounce, the margin of reporting error could be up to 56 g (2 oz).

A strength of our study is that a systematic and comprehensive review process, devised with an experienced librarian, reported in line with PRISMA guidelines, was followed for this review. Two reviewers independently assessed eligibility of the titles, abstracts and full-text studies. We were able to conduct a meta-analysis of a significant number of studies with a large pooled sample size. Studies only including clinical populations, for example, mental or physical illnesses were excluded. We did this to ensure that our results were generalizable to the general population. Future systematic reviews can establish if the findings are similar in clinical subgroups.

However, there are some potential limitations of our study. The search terms were broad, and it is possible we have missed some potentially eligible studies. We also excluded studies that categorized births into three or less groups. The studies are heterogeneous in terms of size, countries, ethnicities, age groups, methodology (e.g. data collection methods, gold standard used), and reporting of statistical analysis. However, we performed sensitivity analyses to assess the influence of several potential influences on results, for example, imperial *v.* metric measurement, sample size, time since birth, first born *v.* subsequent birth, self *v.* parental recall, and found that there was no statistically significant influence on results. We also assessed the effect of the two largest studies: removing them increased the summary estimate from 1.4 g to 5.8 g, but neither of these are clinically significant. A further limitation is that the majority of studies were small, and the overall results are predominantly affected by a few large studies (in qualitative analysis[23,47,48], in meta-analysis[6,14,42,43]). However, the smaller studies had similar findings in qualitative review and meta-analysis.

Any validation study is limited by the data available: here, we required both the availability of a historical record, and an individual's recall. Clinical records may not be accessible in some countries, accurate data may not be recorded particularly in home births. Recovery of recorded birth weights could be as

https://doi.org/10.1017/S2040174416000581 Published online by Cambridge University Press

low as 10%. Historical records require transcription from hand-written ledgers for electronic analyses. Birth certificates include birth weights in some countries (e.g. United States) but not all. Recall rates – where reported – were variable, for example, self-recall 24[12] or 46%.[37] This may vary for several reasons, for example, by country: in Africa up to 25% could not recall birth weight;[46] due to maternal of fetal factors such as maternal education;[47] or due to neonatal complications.[26] Furthermore, there are many methods of reporting the agreement between two measures.

We report correlation and mean difference, but acknowledge that overall correlation coefficient is limited as a measure of agreement: it measures the strength of the relationship between two variables, not the agreement between them; it is unaffected by the scale of measurement (e.g. grams or kilograms); it depends of the range of the measurements; it may mask variability within subgroups, or in certain parts of the distribution.[12,50] The Pearson correlation coefficient is, however, required to correctly estimate the variance of the mean difference, so we would suggest that authors of future studies include this along with other measures of agreement.

We did not assess risk of bias using formal tools: there is currently no consensus on the best method of quality assessment for observational studies. The major source of potential bias was whether the individual had access to the recorded birth weight: for example in Catov *et al.*[13], the mother brought in the birth certificate at the time of interview, which was used as the record of actual birth weight. However, the results were similar in studies where there was no access to the recorded birth weight. Some studies suggest that recall may be more accurate within some ethnic, socioeconomic or clinical subgroups.[6,12] We did not extract data relating to this, and many studies did not report these data.

Birth weight from historical records has been used in many epidemiological studies, particularly relating to the Developmental Origins of Health and Disease.[1,2] It is debated whether recalled birth weight is sufficient to explore the influence of early life factors as part of life course epidemiology. However, it is still widely used, and the findings from this systematic review and meta-analysis suggest that recalled birth weight can be reliability used as an estimate of actual birth weight, where birth records are not available, for example as a risk factor for later disease.[1,2] Recalled birth weight also appears valid in low birth weight and preterm births, as part of population studies, but future studies should explore whether there are different rates of recall in clinical subgroups. There is insufficient evidence to confidently extrapolate this finding to low income countries, and future studies should explore whether the reported recall of higher birth weight in low and middle income countries is replicated, and explore potential reasons for this.

Conclusion

This systematic review and meta-analysis suggests that where birth weight is recalled, it can confidently be used as a reliable estimate of actual birth weight, particularly in high income countries.

Acknowledgments

Thanks to Jillian Hosie for administrative support in sending letters to request missing data and Sheila Fisken for advice on search strategy. Thanks to Sou and Tehranifar for replying to request for additional information.

Financial Support

This work did not receive external funding. The University of Edinburgh Centre for Cognitive Aging and Cognitive Epidemiology (S.D.S.) is part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the Biotechnology and Biological Sciences Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Medical Research Council; and British Heart Foundation (R.M.R.) and Tommy's (R.M.R.) is gratefully acknowledged. G.D. is funded by the Medical Research Council (MC_UU_12017-13).

Conflicts of Interest

None.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S2040174416000581

References

1. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *J Epidemiol Community Health*. 2003; 57, 778–783.
2. Baker JL, Olsen LW, Sørensen TI. Weight at birth and all-cause mortality in adulthood. *Epidemiology*. 2008; 19, 197–203.
3. Leon DA, Lithell HO, Vågerö D, *et al.* Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15 000 Swedish men and women born 1915-29. *BMJ*. 1997; 315, 396–400.
4. Rich-Edwards JW, Stampfer MJ, Manson JE, *et al.* Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. *BMJ*. 1997; 315, 396–400.
5. Casey R, Rieckhoff M, Beebe SA. The accuracy of maternal report. *Clin Pediatr (Phila)*. 1992; 31, 200–205.
6. Tate AR, Dezateux C, Cole TJ, *et al.* Factors affecting a mother's recall of her baby's birth weight. *Int J Epidemiol.* 2005; 34, 688–695.
7. Stroup DF, Berlin JA, Morton SC, *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA*. 2000; 283, 2008–2012.
8. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009; 339, b2535.
9. Sou SC, Chen WJ, Hsieh WS, Jeng SF. Severe obstetric complications and birth characteristics in preterm or term delivery were accurately recalled by mothers. *J Clin Epidemiol*. 2006; 59, 429–435.
10. Tehranifar P, Liao Y, Flom JD, Terry MB. Validity of self-reported birth weight by adult women: sociodemographic

influences and implications for life-course studies. *Am J Epidemiol*. 2009; 170, 910–917.
11. O'Sullivan JJ, Pearce MS, Parker L. Parental recall of birth weight: how accurate is it? *Arch Dis Child*. 2000; 82, 202–203.
12. Andersson SW, Niklasson A, Lapidus L, *et al*. Poor agreement between self-reported birth weight and birth weight from original records in adult women. *Am J Epidemiol*. 2000; 152, 609–616.
13. Catov JM, Newman AB, Kelsey SF, *et al*. Accuracy and reliability of maternal recall of infant birth weight among older women. *Ann Epidemiol*. 2006; 16, 429–431.
14. Jaspers M, de Meer G, Verhulst FC, Ormel J, Reijneveld SA. Limited validity of parental recall on pregnancy, birth, and early childhood at child age 10 years. *J Clin Epidemiol*. 2010; 63, 185–191.
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986; 7, 177–188.
16. Pyles MK, Stolz HR, Macfarlene JW. The accuracy of mothers' reports on birth and developmental data. *Child Dev*. 1935; 6, 165–176.
17. Donoghue EC, Shakespeare RA. The Reliability of peadiatric case-history milestones. *Dev Med Child Neurol*. 1967; 9, 64–69.
18. Bailit HL. Factors affecting mother's estimations of their children's birth weights. *Public Heal Rep*. 1968; 83, 1011–1014.
19. Porteous JR, Meskin LH, Proshek JM, Ten Bensel RW. Epidemiologic assessment of birth weight through maternal recall. *J Dent Res*. 1972; 52, 397.
20. Hoekelman RA, Kelly J, Zimmer AW. The reliability of maternal recall. Mother's remembrance of their infant's health and illness. *Clin Pediatr (Phila)*. 1976; 15, 261–265.
21. Axelsson G, Rylander R. Validation of questionnaire reported miscarriage, malformation and birth weight. *Int J Epidemiol*. 1984; 13, 94–98.
22. Oates RK, Forrest D. Reliability of mothers' reports of birth data. *Aust Pediatr J*. 1984; 20, 185–186.
23. Victora CG, Barros FC, Martines JC, Béria JU, Vaughan JP. AS MÃES LEMBRAM O PESO AO NASCER DE SEUS FILHOS? INTRODUÇÃO. *Rev Saude Publica*. 1985; 19, 195–200.
24. Eaton-Evans J, Dugdale AE. Recall by mothers of the birth weights and feeding of their children. *Hum Nutr Appl Nutr*. 1986; 40, 171–175.
25. Burns TL, Moll PP, Rost CA, Lauer RM. Mothers remember birthweights of adolescent children: The muscatine ponderosity family study. *Int J Epidemiol*. 1987; 16, 550–555.
26. Seidman DS, Slater PE, Ever-Hadani P, Gale R. Accuracy of mothers' recall of birthweight and gestational age. *BJOG*. 1987; 94, 731–735.
27. Wilcox WD, Gold BD, Tuboku-Metzger AJ. Maternal recall of infant birth weight. *Clin Pediatr (Phila)*. 1991; 30, 509–510.
28. Lumey LH, Stein AD, Ravelli CJ. Maternal recall of birthweights of adult children: validation by hospital and well baby clinic records. *Int J Epidemiol*. 1994; 23, 1006–1012.
29. Pless CR, Pless IB. How well they remember. *Arch Pediatr Adolesc Med*. 1995; 149, 553–558.
30. Troy LM, Michels KB, Hunter DJ, *et al*. Self-reported birthweight and history of having been bBreastfed among younger women: an assessment of validity. *Int J Epidemiol*. 1996; 25, 122–127.
31. Olson JE, Shu XO, Ross Ja, Pendergrass T, Robison LL. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. *Am J Epidemiol*. 1997; 145, 58–67.
32. Gaskin P, Walker SP, Forrester TE, Grantham-McGregor SM. The validity of recalled birthweight in developing countries. *Am J Public Health*. 1997; 87, 114.
33. Sanderson M, Williams MA, Malone KE, *et al*. Perinatal factors and risk of breast cancer. *Epidemiology*. 1996; 7, 34–37.
34. Lederman SA, Paxton A. Maternal reporting of prepregnancy weight and birth outcome: consistency and completeness compared with the clinical record. *Matern Child Health J*. 1998; 2, 123–126.
35. Tomeo CA, Rich-Edwards JW, Michels KB, *et al*. Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology*. 1999; 10, 774–777.
36. Gofin R, Neumark YD, Adler B. Birthweight recall by mothers of Israeli children. *Public Health*. 2000; 114, 161–163.
37. Kemp M, Gunnell D, Maynard M, Davey Smith G, Frankel S. How accurate is self reported birth weight among elderly? *J Epidemiol Community Health*. 2000; 54, 639–640.
38. Diaz JF, GIlman RH, Cabrera L. Maternal reporting of birthweight in a Peruvian shanty town. *Am J Public Health*. 2000; 224, 871–872.
39. Walton KA, Murray LJ, Gallagher AM, *et al*. Parental recall of birthweight: a good proxy for recorded birthweight? *Eur J Epidemiol*. 2000; 16, 793–796.
40. Allen DS, Ellison GTH, Dos Santos Silva I, De Stavola BL, Fentiman IS. Determinants of the availability and accuracy of self-reported birth weight in middle-aged and elderly women. *Am J Epidemiol*. 2002; 155, 379–384.
41. Lucia VC, Luo Z, Gardiner JC, Paneth N, Breslau N. Reports of birthweight by adolescents and their mothers: comparing accuracy and identifying correlates. *Paediatr Perinat Epidemiol*. 2006; 20, 520–527.
42. Araujo CLP, Dutra CLC, Hallal PC. Validity of maternal report on birth weight 11 years after delivery: the 1993 Pelotas birth cohort study, Rio Grande do Sul State, Brazil. *Cad Saude Publica*. 2007; 23, 2421–2427.
43. Adegboye ARA, Heitmann BL. Accuracy and correlates of maternal recall of birthweight and gestational age. *BJOG*. 2008; 115, 886–893.
44. Wodskou PM, Hundrup YA, Obel EB, Jørgensen T. Validity of self-reported birthweight among middle-aged and elderly women in the Danish nurse cohort study. *Acta Obstet Gynecol Scand*. 2010; 89, 1134–1139.
45. Boeke CE, Marin C, Oliveros H, *et al*. Validity of maternal birth weight recall among Colombian children. *Matern Child Health J*. 2012; 16, 753–759.
46. Lule SA, Webb EL, Ndibazza J, *et al*. Maternal recall of birthweight and birth size in Entebbe, Uganda. *Trop Med Int Health*. 2012; 17, 1465–1469.
47. Gayle HD, Yip R, Frank MJ, *et al*. Validation of maternally reported birth weights among 46,637 Tennessee WIC program participants. *Public Health Rep*. 2012; 103, 143–147.
48. Bat-Erdene U, Metcalfe A, McDonald SW, Tough SC. Validation of Canadian mothers' recall of events in labour and delivery with electronic health records. *BMC Pregnancy Childbirth*. 2013; 13(Suppl 1), S3.
49. Troude P, L'Hélias LF, Raison-Boulley AM, *et al*. Perinatal factors reported by mothers: do they agree with medical records? *Eur J Epidemiol*. 2008; 23, 557–564.
50. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1, 307–310.