Weight estimation among multi-racial/ethnic infants and children aged 0–5·9 years in the USA: simple tools for a critical measure

Yeyi Zhu[1,2,3,]*, Ladia M Hernandez[3], Yongquan Dong[3], John H Himes[4], Laura E Caulfield[5], Jean M Kerver[6], Lenore Arab[7], Paula Voss[8], Steven Hirschfeld[2,9] and Michele R Forman[3,10]

[1]Kaiser Permanente Northern California Division of Research, 2000 Broadway, Oakland, CA 94612, USA: [2]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA: [3]Department of Nutritional Sciences, University of Texas at Austin, Austin, TX, USA: [4]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA: [5]Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA: [6]Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA: [7]David Geffen School of Medicine, University of California, Los Angeles, CA, USA: [8]Department of Pediatrics, University of California, Irvine, CA, USA: [9]National Institute on Deafness and other Communications Disorders, Bethesda, MD, USA: [10]Department of Nutrition Science, College of Health and Human Science, Purdue University, West Lafayette, IL, USA

Abstract

Objective: In resource-constrained facilities or during resuscitation, immediate paediatric weight estimation remains a fundamental challenge. We aimed to develop and validate weight estimation models based on ulna length and forearm width and circumference measured by simple and portable tools; and to compare them against previous methods (advanced paediatric life support (APLS), Theron and Traub–Johnson formulas).

Design: Cross-sectional analysis of anthropometric measurements. Four ulna- and forearm-based weight estimation models were developed in the training set (*n* 1016). Assessment of bias, precision and accuracy was examined in the validation set (*n* 457).

Setting: National Children's Study-Formative Research in Anthropometry (2011–2012).

Subjects: Multi-racial/ethnic infants and children aged <6 years (*n* 1473).

Results: Developed Models 1–4 had high predictive precision ($R^2 = 0.91–0.97$). Mean percentage errors between predicted and measured weight were significantly smaller across the developed models (0·1–0·7 %) *v.* the APLS, Theron and Traub–Johnson formulas (−1·7, 9·2 and −4·9 %, respectively). Root-mean-squared percentage error was overall smaller among Models 1–4 *v.* the three existing methods (range = 7·5–8·7 *v.* 9·8–13·3 %). Further, Models 1–4 were within 10 and 20 % of actual weight in 72–87 and 95–99 % of the weight estimations, respectively, which outperformed any of the three existing methods.

Conclusions: Ulna length, forearm width and forearm circumference by simple and portable tools could serve as valid and reliable surrogate measures of weight among infants and children aged <6 years with improved precision over the existing age- or length-based methods. Further validation of these models in physically impaired or non-ambulatory children is warranted.

Measurement of weight is one of the most fundamental anthropometric measures and an essential indictor for growth and nutritional status in clinical care and paediatric research. Weight is conventionally determined by a mechanical or electronic scale, if available. However, immediate and accurate weight measurement remains a fundamental challenge in situations where the child is immobilized due to critical illness or acute injury in emergency settings. Indeed, weight is a vital measurement performed in paediatric emergency departments and is critical for diagnostic and therapeutic decisions, such as estimating energy requirements and calculating

Corresponding author: Email yeyi.zhu@kp.org

individualized medication dosage, fluid administration and device sizes. Thus, failure to accurately estimate paediatric weight could comprise the quality of paediatric care.

Although parental recall or weight estimation by caregivers may be available in certain circumstances, the accuracy varies widely and may lack consistency in different populations[1-7]. Therefore, various weight estimation methods have been developed, mostly based on a child's age, length or both. Overall, length- or length- and age-based methods have greater accuracy than solely age-based ones[8,9]; however, accurate measurement of recumbent length, particularly in infants and young children, has its challenges[10]. Moreover, most previous weight estimation methods tend to under- or overestimate weight in children at the extremes of the weight distribution[11]. Given the childhood obesity epidemic in high-income countries and the prevalence of both underweight and overweight/obese children in low- and middle-income countries[12,13], weight estimation strategies which accommodate children across weight categories with consistent, improved precision over the existing methods are warranted.

In the present study, we examined the accuracy and reliability of ulna length, a previously validated surrogate for paediatric length/height[14], and forearm measurements (width and circumference) measured by simple and portable tools as surrogates of paediatric weight in a multiracial/ethnic population of infants and children aged <6 years in the USA. Further, to assess the performance of these ulna- and forearm-based weight estimation models, we compared them with several existing age- or length-based models (i.e. advanced paediatric life support (APLS), Theron and Traub–Johnson formulas; Table 1)[15-17].

Methods

Study design and population

The study was a cross-sectional assessment of anthropometric status of infants/children aged <6 years across eight study centres in the USA (2011–2012). The detailed design

of the study has been described previously[14,18]. Briefly, mother–offspring dyads were recruited at daycare centres, churches, clinics and community centres (*n* 1634). Eligibility criteria were: mothers aged 18–49 years and non-institutionalized; and offspring who were aged 0–5·9 years, healthy, had not suffered from any illness associated with weight loss during the past week, and were afebrile at the time of study visit. If more than one infant/child of the mother was recruited, the youngest singleton was included in the current analysis to reduce the cluster effect within the same family (*n* 1560). The analysis included infants/children with at least one anthropometric measurement (*n* 1473).

Data collection

Child's age, sex and race/ethnicity were reported by the mother using an interviewer-administered questionnaire. Anthropometric measurements were obtained by data collection teams each composed of two trained researchers (one measurer, one recorder). Following standard anthropometric protocols[19], weight was measured to the nearest 0·01 kg in infants wearing a dry diaper or in children wearing underpants on an electronic scale (SECA, Germany), calibrated daily using a Troemner® weight. Recumbent length and standing height were measured to the nearest 0·1 cm using an infantometer and a portable stadiometer (SECA, Germany) in infants/children aged 0–1·9 and 2–5·9 years, respectively.

All ulna measurements were obtained on the right arm. After marking the two end points of the ulna (i.e. the styloid and olecranon processes), ulna length was measured to the nearest 0·1 cm using a calliper (Rosscraft Innovations Inc., Canada) while the right arm was placed in a horizontal plane with the elbow flexed ~90° (see online supplementary material, Supplemental Fig. 1A). Forearm width was measured to the nearest 0·1 cm using a graph paper grid which can be printed on a regular letter-size paper by: (i) having the participant place his/her arm on a table or a thin rigid board (such as a clipboard); (ii) having the right arm straightened and pointing outward from the body with palm down and lateral aspect of the forearm aligned along the zero vertical axis of the grid; (iii) marking two points at the maximal width of the forearm on the grid; and (iv) reading the maximal width of the forearm to the nearest 0·1 cm according to the uniform dimensions on the grid (see online supplementary material, Supplemental Fig. 1B). Of note, the grid was coloured across rows/units of ten boxes to facilitate reading the measurements. Forearm circumference was measured to the nearest 0·1 cm using an insertion tape (ShorrTape©, USA) on the forearm by: (i) having the right elbow extended and the forearm positioned so that it is free-standing (not resting on the table or body); (ii) having the tape measure perpendicular to the long axis of the forearm; and (iii) measuring the maximal forearm

Table 1 Previous age- or length-based methods for weight estimation in children

Method	Age limit (years)	Equation
APLS formula[15,41]	1–5	Weight (kg) = 2 × [age (years) + 4]
	6–9	Weight (kg) = 3 × [age (years) + 7]
Theron formula[16]	1–10	Weight (kg) = exp[0·175571 × age (years) + 2·197099]
Traub-Johnson formula[17]	1–18	Weight (kg) = 2·05 × exp[0·02 × length or height (cm)]

APLS, advanced paediatric life support.

circumference with the tape measure (see online supplementary material, Supplemental Fig. 1C).

Each measurement was taken in duplicate. The mean value was calculated if the two initial measurements agreed within 0·2 kg for weight or 0·2 cm for length, height, and ulna and forearm measurements. Otherwise, an additional measurement was obtained and the mean of the two closest recordings was used. To determine the intra- and inter-observer reliability, replicate measures were taken by reversing staff's positions as measurer and recorder in an approximately 10 % random sub-sample (n 124).

Statistical analysis

Data pre-processing approaches were reported previously[20]. Based on the point biserial model for correlations, the total sample size of 1473 was sufficient to detect an effect size as small as $r = 0·07$ between an ulna or forearm measurement and weight at 80 % power with a two-tailed significance level of $\alpha = 0·05$. The total sample was randomly split 2:1 into a training set (n 1016) and validation set (n 457). Comparison of subject characteristics and anthropometrics between the two sets was tested by the Student's t test for continuous variables or the χ^2 test for categorical variables. Intra- and inter-observer reliability of each anthropometric measure was estimated by computing CV and intraclass correlation coefficients (ICC) using a one-way random model and absolute agreement type[21] in the random sub-sample of 124 infants/children.

Prediction equations for weight were developed in the training set using multivariable mixed-effects linear regression analysis with study centre as a random effect. Initially, parameters for stature (length/height or ulna length) and body size (forearm width or circumference) were included as predictors. Given significant age, sex and racial/ethnic variation in anthropometrics (see online supplementary material, Supplemental Table 1), we included these factors as potential predictors. Notably, racial/ethnic variation was parameterized as a dichotomous variable (i.e. Hispanic or not) given the over-sampling of Hispanics in our study population. Also, we included a quadratic term for forearm width or circumference in all models given the non-linear associations observed between weight and forearm width or circumference. Final models were reduced by stepwise elimination using entry ($P = 0·10$) and removal ($P = 0·05$) criteria. The marginal R^2 proposed by Nakagawa and Schielzeth was calculated to represent the proportion of variance explained by fixed effects[22]. Standard error of estimate was computed for each equation.

In the validation set, mean percentage error (MPE), a measure of the overall bias estimate of each model, was calculated as: $100 \times$ (predicted weight – measured weight)/measured weight. Root-mean-squared percentage error (RMSPE), a measure of precision estimate, was calculated by taking the square root of the average squared percentage error. Percentages of weight estimates falling within 10 and 20 % limits of deviation from actual weight were calculated to assess the predictive accuracy. Comparison of the aforementioned estimates between existing methods and newly developed models were assessed using paired t tests with Bonferroni–Holm adjustment[23] by: weight strata (<10, 10–19·9 and ≥20 kg) for all; weight-for-length Z-score (WLZ) percentile categories (i.e. underweight/normal (WLZ < 85th percentile) and overweight/obese (WLZ ≥ 85th percentile)) among infants aged <2 years; and BMI-for-age Z-score (BMIZ) percentile categories (i.e. underweight/normal (BMIZ < 85th percentile) and overweight/obese (BMIZ ≥ 85th percentile)) among children aged 2–5·9 years. As recommended by the Centers for Disease Control and Prevention, BMI is used to screen for overweight/obesity in children ≥2 years old[24]. Therefore, infants aged <2 years were grouped separately according to WLZ percentiles derived from the WHO Child Growth Standards[25].

Further, Bland–Altman plots[26] were constructed to assess the agreement between the measured and predicted weight by our models and existing ones[15–17]. The limits of agreement were defined as the mean difference between the predicted and measured weight ±1·96 SD. We constructed Bland–Altman plots on the original scale (i.e. kilograms) given the narrow age and weight range of the study. This approach however, compared with log-transformation of the data, also allows direct evaluation on the original scale of the agreement between predicted and measured weight, which could facilitate interpretation within the context of real settings.

All analyses were conducted with the statistical software package IBM SPSS Statistics version 21 and R software version 3·3. Statistical significance was set at a two-tailed $P < 0·05$.

Results

Among the 1016 infants/children in the training set, 52·3 % were boys; the overall mean age was 1·9 years; and the ethnic distribution was 45·6 % Hispanic, 25·5 % non-Hispanic Black, 20·5 % non-Hispanic White and 8·4 % Other groups (Table 2). The validation set did not differ from the training set by demographic characteristics or anthropometric measures. All anthropometric measures including ulna and forearm measurements had high intra- and inter-observer reliability overall, with CV ranging from 0·08 to 2·16 % and ICC ranging from 0·952 to 1·000 (see online supplementary material, Supplemental Table 2). Weight measured by calibrated scale had the highest intra-observer reliability with the smallest CV and greatest ICC, followed by height, length, forearm circumference, ulna length and forearm width. Likewise, weight had the

Table 2 Subject characteristics and child anthropometrics in the training and validation sets of multi-racial/ethnic infants and children aged <6 years; National Children's Study-Formative Research in Anthropometry, USA (2011–2012)

	Training set (*n* 1016)		Validation set (*n* 457)		
	n or Mean	% or SD	*n* or Mean	% or SD	*P* value*
Sex, *n* and %					0·65
Boys	531	52·3	233	51·0	
Girls	485	47·7	224	49·0	
Race/ethnicity, *n* and %					0·74
Non-Hispanic White	208	20·5	88	19·3	
Hispanic	462	45·6	200	43·9	
Non-Hispanic Black	258	25·5	126	27·6	
Other	88	8·4	43	9·2	
Age (years), mean and SD	1·9	1·7	1·9	1·7	0·27
Anthropometrics, mean and SD					
Weight (kg)	11·5	5·3	11·8	5·0	0·37
Recumbent length (cm)†	70·7	11·9	71·9	12·1	0·15
Standing height (cm)†	100·7	8·8	100·6	8·6	0·96
Ulna length (cm)	12·7	2·9	12·8	2·9	0·48
Forearm width (cm)	5·3	0·7	5·3	0·6	0·99
Forearm circumference (cm)	15·0	2·2	15·1	2·0	0·42

*Obtained by Student's *t* test for continuous variables and the χ^2 test for categorical variables.
†Recumbent length and standing height were measured among infants and children aged 0–1·9 and 2–5·9 years, respectively.

highest whereas forearm width had the lowest inter-observer reliability, respectively.

In total, four weight estimation models were empirically derived as listed in Table 3. Of note, age and sex were not included in Models 1 and 2 due to the insignificant contribution to the final models according to the stepwise elimination criteria mentioned above. Overall, models using total body length/height as a predictor (Models 1 and 2) and models using ulna length as a surrogate for length/height (Models 3 and 4) had comparable predictive accuracy, regardless of the surrogate for body size (forearm width or circumference). Further, among the two models using ulna length as a surrogate for length/height, the one using forearm circumference as a surrogate for body size (Model 4) had slightly greater predictive accuracy than the one using forearm width (Model 3).

Overall, compared with the three existing formulas, the performance of Models 1–4 did not differ appreciably between one another and was superior to the APLS, Theron and Traub–Johnson formulas (Table 4). Across the weight strata, the MPE were significantly smaller across Models 1–4 compared with the existing formulas except that the Traub–Johnson did not vary from Models 2–4 among infants/children weighing <10 kg (1·2 *v.* −0·5 to 2·4%); and that the Theron formula did not vary from Models 1–4 among infants/children weighing ≥20 kg (−6·2 *v.* −4·8 to −8·1%). Among infants aged <2 years with WLZ < 85th percentile, the MPE were 0·2 to 1·4% across Models 1–4, which were significantly smaller than for the APLS (6·1%), Theron (13·3%) and Traub-Johnson (−2·4%) formulas. Among infants aged <2 years with WLZ ≥ 85th percentile, all models tended to underestimate weight except the Theron formula (5·2%); Models 3 and 4 slightly underestimated weight by −2·4 to −1·7%, followed by Models 1 and 2 (−6·2 to −4·3%) and the APLS formula

(−4·3%), whereas the Traub–Johnson formula had the greatest MPE (−14·3%). For underweight/normal-weight children aged 2–5·9 years, Model 4 and the APLS formula slightly overestimated weight by 2·3 and 0·8%, respectively, whereas the Theron formula had the largest MPE (14·3%). Among overweight/obese children aged 2–5·9 years, all models tended to underestimate paediatric weight; however, Models 2 and 4 yielded the smallest MPE (−5·2% and −4·1%) and the APLS formula yielded the greatest (−18·0%). Consistently, the measure of precision as indicated by RMSPE was overall smaller among Models 1–4 compared with the three existing methods (i.e. range = 7·5–8·7 *v.* 9·8–13·3%; Table 4). The differences in RMSPE across models were more pronounced at weight extremes, i.e. among children weighing <10 kg or ≥20 kg or overweight/obese infants or children. Further, estimates of accuracy as indicated by percentage of agreement within 10 and 20%% limits of deviation from actual weight illustrated that the predictive accuracy was greater across Models 1–4 compared with the three existing methods (Table 5). Specifically, Models 1–4 were overall within 10 and 20% of actual weight in 72·2–86·9 and 95·2–98·5% of the weight estimations, respectively, which outperformed any of the other existing methods (56·5–68·6 and 74·5–83·0% of weight estimations within 10 and 20% of actual weight, respectively).

Overall, the Bland–Altman plots illustrated no obviously biased patterns of paediatric weight estimation using Models 1–4 (mean difference range = −0·012 to 0·002 kg), especially among infants (corresponding to the small values on the *x*-axis; Fig. 1). In contrast, the APLS, Theron and Traub–Johnson formulas tended to underestimate weight (mean difference range = −0·602 to −0·962 kg) as the mean values of weight increased. In addition, the limits of agreement were narrower for Models 1–4 compared

Table 3 Regression equations* to estimate weight in infants and children aged 0–5·9 years developed in the training set (n 1016) of multi-racial/ethnic infants and children aged <6 years; National Children's Study-Formative Research in Anthropometry, USA (2011–2012)

	Equation	$R^2_{marginal}$	SEE (kg)
Model 1	Weight (kg) $= 15.89 + 0.35 \times$ Hispanic $+ 0.25 \times L - 10.63 \times FW + 1.12 \times FW^2$	0.94	1.23
Model 2	Weight (kg) $= 6.91 + 0.40 \times$ Hispanic $+ 0.21 \times L - 2.51 \times FC + 0.10 \times FC^2$	0.97	0.96
Model 3	Weight (kg) $= 19.45 + 1.15 \times A + 0.22 \times$ Hispanic $+ 0.76 \times UL - 9.21 \times FW + 1.03 \times FW^2$	0.91	1.46
Model 4	Weight (kg) $= 6.25 + 1.04 \times A + 0.37 \times$ Hispanic $+ 0.59 \times UL - 1.65 \times FC + 0.09 \times FC^2$	0.96	1.09

$R^2_{marginal}$, coefficient of determination for fixed effects; SEE, standard error of estimate (kg); L, length/height (cm); FW, forearm width (cm); FC, forearm circumference (cm); A, age (years); UL, ulna length (cm).
*Equations were obtained from mixed-effects linear regression analysis using study centre as a random effect.

Table 4 Mean percentage error (MPE)* between predicted and measured weight and root-mean-squared percentage error (RMSPE) by weight, weight-for-length percentile and BMI-for-age percentile categories in the validation set (n 457) of multi-racial/ethnic infants and children aged <6 years; National Children's Study-Formative Research in Anthropometry, USA (2011–2012)

	Model 1		Model 2		Model 3		Model 4		APLS formula		Theron formula		Traub–Johnson formula	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total (n 457)														
MPE	0.2^a	9.6	0.1^a	7.5	0.7^a	11.2	$0.2^{a,b}$	8.7	-1.7^b	13.2	9.2^c	15.1	-4.9^d	8.5
RMSPE	8.6		7.5		8.7		8.3		13.3		17.7		10.5	
Weight category														
<10 kg (n 190)														
MPE	-0.8^a	9.0	$-0.5^{a,b}$	8.9	2.4^b	11.9	$0.7^{a,b}$	10.9	14.6^c	7.9	22.5^d	8.4	1.2^b	6.4
RMSPE	8.0		7.9		9.6		9.9		16.6		24.0		8.4	
10–19.9 kg (n 244)														
MPE	1.7^a	9.4	1.2^a	6.1	0.5^a	10.6	0.2^a	6.9	-1.8^b	11.0	9.0^c	14.4	-4.9^d	7.9
RMSPE	9.2		6.2		7.6		6.8		11.1		17.0		12.3	
≥20 kg (n 23)														
MPE	-8.1^a	9.9	-6.1^a	4.3	-7.0^a	10.3	-4.8^a	0.2	-22.4^b	9.4	-6.2^a	13.1	-13.5^c	8.9
RMSPE	9.7		7.4		9.3		6.3		24.2		14.3		16.1	
Infants aged <2 years														
Underweight/normal weight† (n 199)														
MPE	1.3^a	7.9	1.0^a	7.9	0.2^a	10.4	1.4^a	11.5	6.1^b	9.6	13.3^c	10.4	-2.4^d	5.7
RMSPE	7.1		5.4		7.3		6.2		11.0		20.1		8.6	
Overweight/obese† (n 69)														
MPE	-6.2^a	7.2	$-4.3^{a,b}$	5.8	-2.4^a	8.8	-1.7^a	8.7	-4.3^b	11.5	5.2^c	12.4	-14.3^d	5.7
RMSPE	8.8		6.9		8.6		7.4		20.5		12.9		16.7	
Children aged 2–5.9 years														
Underweight/normal weight† (n 144)														
MPE	4.6^a	7.8	$3.2^{a,b}$	4.3	$3.6^{a,b}$	9.6	2.3^c	5.8	0.8^c	11.0	14.3^d	14.1	-1.6^e	6.4
RMSPE	8.0		8.0		9.7		7.4		11.3		16.8		6.8	
Overweight/obese† (n 45)														
MPE	-8.6^a	8.2	-5.2^b	4.5	-8.1^a	9.8	-4.1^b	6.2	-18.0^c	9.9	$-6.5^{a,b}$	11.3	-15.8^c	5.5
RMSPE	9.1		8.9		9.7		8.3		14.0		14.2		15.7	

APLS, advanced paediatric life support.
[a,b,c,d]Mean values in a row with unlike superscript letters were significantly different ($P < 0.05$) using the paired t test with Bonferroni–Holm adjustment for pairwise comparisons.
*MPE was calculated as: $100 \times$ (predicted weight − measured weight)/measured weight.
†Underweight/normal and overweight/obese were defined as <85th percentile and ≥85th percentile for weight-for-length among infants aged <2 years and for BMI-for-age for among children aged 2–5.9 years, respectively.

with the existing formulas, with the APLS having the widest range (−5·10 to 3·90 kg).

Discussion

In the current study, ulna and forearm measurements obtained by simple, portable and convenient tools (i.e. calliper, paper grid and insertion tape) were accurate and reliable surrogate measures for paediatric weight among healthy infants/children aged <6 years in the USA. The intra- and inter-reliability of ulna and forearm measurements was high and comparable to or better than those reported previously[27–29], suggesting their applicability by trained staff in varied settings including daycare centres, clinics and community centres, as demonstrated in our study. The estimates of predictive bias, precision and accuracy of our empirically derived models were comparable with one another and significantly superior to the three examples of existing age- or length-based formulas, suggesting that they may serve as alternative strategies for paediatric weight estimation when immediate weight measurement is unobtainable or unreliable such as in the emergency room.

Table 5 Predictive accuracy performance* of Models 1–4 and the three existing methods in multi-racial/ethnic infants and children aged <6 years; National Children's Study-Formative Research in Anthropometry, USA (2011–2012)

	Model 1 (%)	Model 2 (%)	Model 3 (%)	Model 4 (%)	APLS formula (%)	Theron formula (%)	Traub–Johnson formula (%)
Total							
≤10 %	79·0	86·9	72·2	79·9	61·6	56·5	68·6
≤20 %	97·8	98·5	95·2	96·5	82·8	74·5	83·0
Weight category							
<10 kg							
≤10 %	81·4	79·8	77·3	76·3	54·8	44·1	65·8
≤20 %	96·8	96·8	92·6	92·6	83·0	70·7	89·2
10–19·9 kg							
≤10 %	79·1	92·2	73·4	85·7	77·8	74·6	87·3
≤20 %	98·9	99·6	96·3	99·2	86·8	79·4	88·9
≥20 kg							
≤10 %	76·5	87·0	76·5	87·0	28·7	47·8	52·2
≤20 %	92·0	100·0	92·0	100·0	56·5	82·6	69·6
Infants aged <2 years							
Underweight/normal weight†							
≤10 %	81·9	94·4	73·6	88·9	51·8	31·7	81·0
≤20 %	95·8	100·0	94·4	99·3	83·8	60·1	85·6
Overweight/obese†							
≤10 %	75·6	88·9	80·0	77·8	22·2	51·1	15·6
≤20 %	93·3	100·0	98·9	100·0	54·4	73·9	72·2
Children aged 2–5·9 years							
Underweight/normal weight†							
≤10 %	88·0	84·4	72·4	73·4	77·0	67·8	81·9
≤20 %	96·9	98·4	96·2	96·2	87·9	80·6	86·6
Overweight/obese†							
≤10	71·8	79·6	75·0	76·8	72·5	77·9	56·4
≤20	95·9	92·9	95·9	96·4	81·4	81·4	83·2

APLS, advanced paediatric life support.
*Values presented are percentages of weight estimates within specified limits of deviation (≤10 and ≤20 %) from actual weight.
†Underweight/normal and overweight/obese were defined as <85th percentile and ≥85th percentile for weight-for-length among infants aged <2 years and for BMI-for-age among children aged 2–5·9 years, respectively.

The high comparability of these four models could provide flexibility and enhance applicability in different settings. In situations where the child's age is unknown, Models 1 and 2 could be utilized for immediate weight estimation, whereas Models 3 and 4 could be utilized when the child's recumbent length or standing height cannot be measured, given measurements of the ulna and forearm are usually not impeded by joint deformity and the ulna is readily accessible even in immobilized patients. Further, Model 3 had the lowest MPE between predicted and measured weight across all models of underweight or normal-weight infants aged <2 years. Taken together, in field settings where a calibrated scale or level floor is unavailable, the ulna and forearm measurements obtained by simple and affordable tools could potentially provide alternative options for paediatric weight estimation, with overall exchangeability and also flexibility in varied settings.

Several strategies for paediatric weight estimation have been developed with varied degrees of applicability in specific paediatric sub-populations. The age-based strategies such as the APLS[15] and Theron[16] formulas have advantages due to their simplicity and lack of additional anthropometric surrogates. However, the APLS formula largely underestimates weight among children weighing more than 20 kg or overweight/obese children by approximately 20 % in our study population, similar to previous observations[10,30]. In contrast, the Theron formula did not vary from our models in terms of predictive accuracy among heavier children, but tended to overestimate weight by 22·5 % among children weighing <10 kg and by 13·3 % among underweight/normal-weight infants aged <2 years. Indeed, the Theron method was developed among a sample of children of Pacific Island and Māori origins in New Zealand, whose overweight/ obesity prevalence was significantly higher than their European counterparts (40–60 *v.* 24 %)[31], potentially limiting their applicability for other paediatric populations. There are several other age-based formulas for paediatric weight estimation, such as the Luscombe formula[32], the finger counting method[33] and the Chinese age–weight rule[34]. As demonstrated in a recent study assessing twenty age-based weight estimation methods, the age-based methods had an overall high rate of critical errors (i.e. percentages of weight estimates falling outside 20 % deviation from actual weight) ranging from 25 to 75 % and were inferior to any length-based method (e.g. Broselow tape, paediatric advanced weight-prediction in the emergency room (PAWPER) tape or the Mercy method)[9].

The length-based strategies such as the Traub–Johnson formula[17] and the Broselow tape[35] could be applied
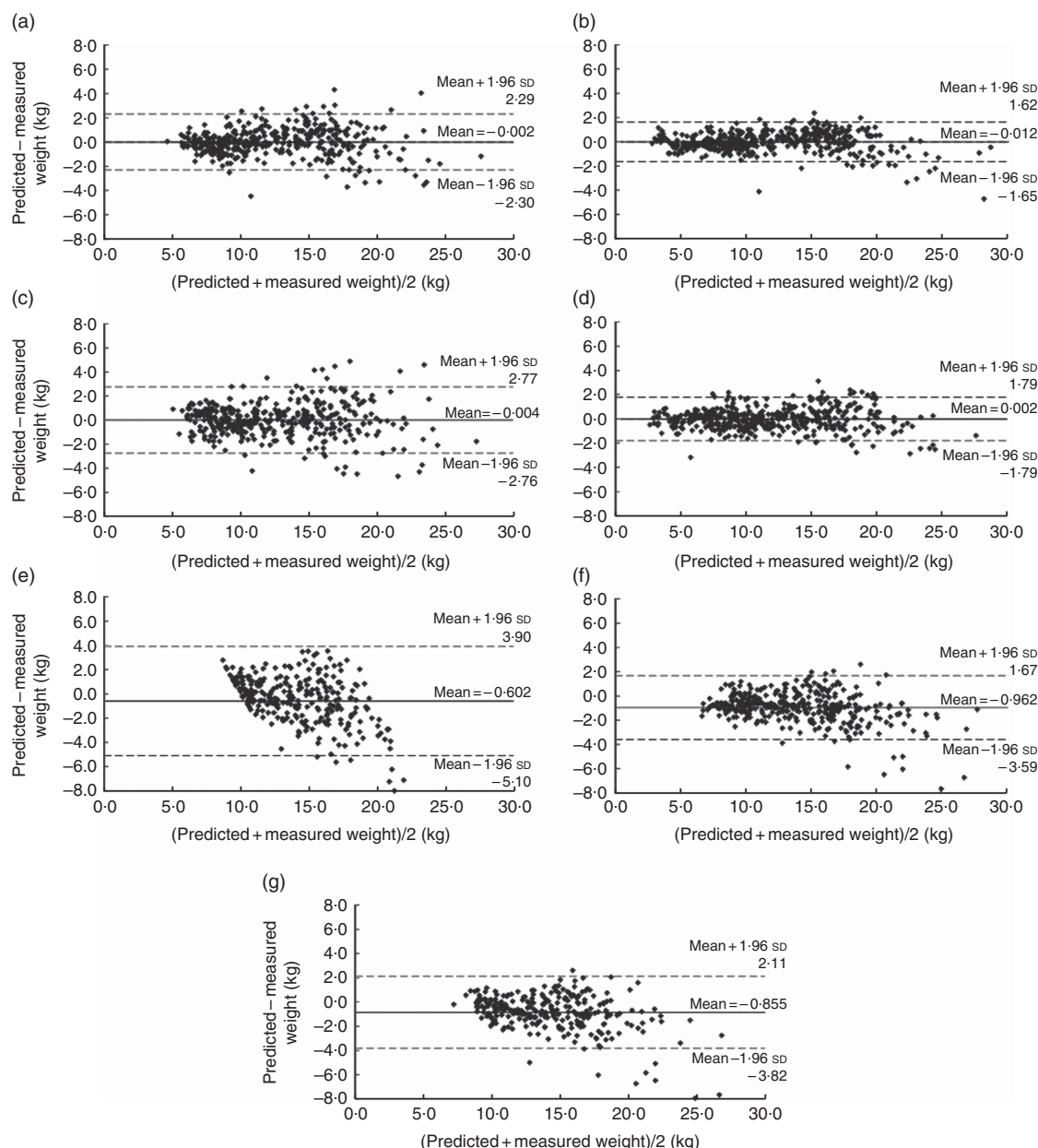
Fig. 1 Bland–Altman plots assessing the relative validity of four weight estimation models (based on ulna length and forearm width and circumference, measured by simple and portable tools) and three previous methods (based on age or length) in predicting weight in multi-racial/ethnic infants and children aged <6 years; National Children's Study-Formative Research in Anthropometry, USA (2011–2012). The difference between measured weight and predicted weight is plotted v. the mean weight from the two methods for: (a) Model 1 (n 376); (b) Model 2 (n 422); (c) Model 3 (n 386); (d) Model 4 (n 431); (e) the advanced paediatric life support (APLS) formula (n 285); (f) the Theron formula (n 285); and (g) the Traub–Johnson formula (n 286) in the validation set. ——— indicates the mean difference (bias) between the predicted and measured weight and – – – – – indicate the 95 % limits of agreement

in situations without knowledge of a child's exact age. Although the measurer can directly read weight from the measuring tape, the Broselow tape is limited to a length range of 46–143 cm. The Traub–Johnson formula had similar prediction accuracy as our Models 2 and 4 among children weighing <10 kg. Nevertheless, its performance was compromised and inferior to our Models 2 and 4 with a bias pattern of underestimation among heavier (≥20 kg) or more obese children. On the other hand, the age- or length-based equations do not take into account the child's body size, which is an important predictor of paediatric weight[11,36]. The Devised Weight Estimation Method[36], a length- and body size-based method, has relatively high prediction accuracy, with MPE between predicted and measured weight ranging from −3·9 to 7·0 % among children weighing <10 to >40 kg. Notably, this method involves a subjective assessment of body size (slim, average or heavy), which may have bias as

evidenced by mean intra- and inter-rater agreement of 86 % (range = 81–94 %) and 78 % (58–93 %), respectively[10]. Similarly, the PAWPER tape involves a two-step process based on supine length and habitus scoring, whereas the accuracy and reliability of the habitus evaluation in different settings remain to be assessed[37].

Among long bone- and/or mid-upper arm circumference (MUAC)-based methods, an MUAC-based formula developed among Hong Kong Chinese children aged 1–11 years outperformed the Broeselow method and the age-based APLS formula in older children, but not among pre-school children under 6 years old[38]. Among predominantly HIV-positive children aged 1·5–12 years in Botswana, over 90 % of the predicted weight fell within 15 % of the actual weight using an MUAC- and tibia or ulna length-based method developed by Wozniak et al.[39,40]. However, due to the limited number of children aged <5 years (*n* 203) and weighing <10 kg (*n* 28), no conclusions can be drawn about these subgroups[39,40]. Further, validity of this method among other paediatric populations remains to be determined. In contrast, the recently developed Mercy method relies on humerus length and MUAC, and has comparable prediction accuracy among children aged 2 months to 16 years to our ulna-/forearm-based models (MPE = – 0·46 *v.* 0·1–0·7 %)[11]. Notably, among children with shoulder/upper arm contractures and/or other physical impairments whose upper arm and total length/height measurements are not feasible, our ulna-/forearm-based models (Models 3 and 4) could serve as alternative strategies for weight estimation. Future studies on other paediatric populations are warranted to further assess the prediction precision of our developed methods in clinical settings.

Certain limitations of the present study should be noted. First, 44·9 % of our study population was of Hispanic origin. Given the limited anthropometric data among Hispanic neonates, infants and young children in the USA, we oversampled Hispanic infants and young children to enrich the limited data on anthropometrics, especially measurements of bone components. The ethnic component in Models 1–4 was dichotomized (i.e. Hispanic or not) given the respective sample size of each ethnic group. Therefore, the study population was not nationally representative which may limit the generalizability of our models. Nevertheless, our models highlight the need for future research to consider and incorporate race/ethnicity in weight prediction strategies among multi-racial/ethnic children. In addition, despite the overall zero bias as shown in the Bland–Altman plots, Models 1 and 3 exhibited some heteroscedasticity in weight estimation at older ages. We oversampled neonates and infants aged <1 years (39 %) to address the data gap given that most previous weight estimation methods are limited to 1 year or above[11]. It is possible that the observed heteroscedasticity could be partially attributable to the insufficient statistical power among older children. Thus, age-specific weight prediction equations based on these surrogate measures merit further investigation. Finally, the impact of human factor and patient factor errors could be significant, especially for methods including any form of anthropometric measurements. Thus, these study findings need to be carefully evaluated during real or simulated emergency care.

Conclusion

In conclusion, ulna and forearm components can serve as accurate and reliable surrogate measures of weight in healthy infants/children aged 0–5·9 years. The developed models for paediatric weight estimation could potentially provide improvement over existing methods, especially among infants. In addition, the use of ulna length as a surrogate for length/height provides an alternative strategy in situations where length/height is not obtainable or unreliable. Further, ulna and forearm measurements can be obtained by simple and portable tools (i.e. calliper, paper grid and tape), which would be valuable in field settings where calibrated equipment (i.e. infantometer, stadiometer or electronic scale) is unavailable due to issues of portability, accessibility and expense. Finally, further evaluation and validation of these developed models are warranted in other paediatric populations, particularly among physically impaired or non-ambulatory children as well as children in resource-limited settings such as in low-income countries or rural areas.

Acknowledgements

to data management and statistical aspects of the work; Y.Z. and M.R.F. had primary responsibility for final content of the manuscript. All authors contributed to manuscript review. All authors read and approved the final manuscript. *Ethics of human subject participation:* This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects/patients were approved by all study centres listed in the online supplementary material, Supplemental Table 3. Written informed consent was obtained from all subjects.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1368980018002549

References

1. Anglemyer BL, Hernandez C, Brice JH *et al.* (2004) The accuracy of visual estimation of body weight in the ED. *Am J Emerg Med* **22**, 526–529.
2. Partridge RL, Abramo TJ, Haggarty KA *et al.* (2009) Analysis of parental and nurse weight estimates of children in the pediatric emergency department. *Pediatr Emerg Care* **25**, 816–818.
3. Harris M, Patterson J & Morse J (1999) Doctors, nurses, and parents are equally poor at estimating pediatric weights. *Pediatr Emerg Care* **15**, 17–18.
4. Lundahl A, Kidwell KM & Nelson TD (2014) Parental underestimates of child weight: a meta-analysis. *Pediatrics* **133**, e689–e703.
5. Rosenberg M, Greenberger S, Rawal A *et al.* (2011) Comparison of Broselow tape measurements versus physician estimations of pediatric weights. *Am J Emerg Med* **29**, 482–488.
6. Williams B, Boyle M & O'Meara P (2010) Can undergraduate paramedic and nursing students accurately estimate patient age and weight? *Prehosp Disaster Med* **25**, 171–177.
7. Rosenberg M, Thundiyil J, Greenberger S *et al.* (2010) 140: Does physician estimates of pediatric patient weights lead to inaccurate medication dosages. *Ann Emerg Med* **56**, 3 Suppl., S47.
8. Georgoulas VG & Wells M (2016) The PAWPER tape and the Mercy method outperform other methods of weight estimation in children at a public hospital in South Africa. *S Afr Med J* **106**, 933–939.
9. Wells M, Goldstein LN & Bentley A (2017) It is time to abandon age-based emergency weight estimation in children! A failed validation of 20 different age-based formulas. *Resuscitation* **116**, 73–83.
10. Black K, Barnett P, Wolfe R *et al.* (2002) Are methods used to estimate weight in children accurate? *Emerg Med (Fremantle)* **14**, 160–165.
11. Abdel-Rahman SM & Ridge AL (2012) An improved pediatric weight estimation strategy. *Open Med Devices J* **4**, 87–97.
12. Ogden CL, Carroll MD, Kit BK *et al.* (2014) Prevalence of childhood and adult obesity in the United States, 2011–2012. *JAMA* **311**, 806–814.
13. Doak CM, Adair LS, Bentley M *et al.* (2005) The dual burden household and the nutrition transition paradox. *Int J Obes (Lond)* **29**, 129–136.
14. Forman MR, Zhu Y, Hernandez LM *et al.* (2014) Arm span and ulnar length are reliable and accurate estimates of recumbent length and height in a multiethnic population of infants and children under 6 years of age. *J Nutr* **144**, 1480–1487.
15. Mackway-Jones K, Molyneux E, Phillips B *et al.* (2005) *Advanced Paediatric Life Support*, 4th ed. London: BMJ Books.
16. Theron L, Adams A, Jansen K *et al.* (2005) Emergency weight estimation in Pacific Island and Maori children who are large-for-age. *Emerg Med Australas* **17**, 238–243.
17. Traub SL & Johnson CE (1980) Comparison of methods of estimating creatinine clearance in children. *Am J Hosp Pharm* **37**, 195–201.
18. Zhu Y, Hernandez LM, Dong Y *et al.* (2015) Longer breastfeeding duration reduces the positive relationships among gestational weight gain, birth weight and childhood anthropometrics. *J Epidemiol Community Health* **69**, 632–638.
19. Lohman TG, Roche AF & Martorell R (1988) *Anthropometric Standardization Reference Manual*. Champaign, IL: Human Kinetic Books.
20. Zhu Y, Hernandez LM, Mueller P *et al.* (2013) Data acquisition and preprocessing in studies on humans: what is not taught in statistics classes? *Am Stat* **67**, 235–241.
21. Landers R (2015) Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *Winnower* **4**, e143518.181744.
22. Nakagawa S & Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* **4**, 133–142.
23. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* **6**, 65–70.
24. Kuczmarski R, Ogden C & Guo S (2002) 2000 CDC growth charts for the United States: methods and development. *Vital Health Stat 11* issue 246, 1–190.
25. World Health Organization (2006) *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development*. Geneva: WHO.
26. Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.
27. Gauld LM, Kappers J, Carlin JB *et al.* (2004) Height prediction from ulna length. *Dev Med Child Neurol* **46**, 475–480.
28. Klipstein-Grobusch K, Georg T & Boeing H (1997) Interviewer variability in anthropometric measurements and estimates of body composition. *Int J Epidemiol* **26**, Suppl. 1, S174–S180.
29. Pappas ND, Watson JT, Erickson JM *et al.* (2013) Reliability and accuracy of templating humeral and ulnar components for total elbow arthroplasty. *Am J Orthop* **42**, 321–323.
30. Loo PY, Chong SL, Lek N *et al.* (2013) Evaluation of three paediatric weight estimation methods in Singapore. *J Paediatr Child Health* **49**, E311–E316.
31. Ministry of Health (2003) *NZ Food NZ Children: Key Results of the 2002 National Children's Nutrition Survey*. Wellington: Ministry of Health.
32. Luscombe MD, Owens BD & Burke D (2011) Weight estimation in paediatrics: a comparison of the APLS formula and the formula 'Weight = 3 (age)+7'. *Emerg Med J* **28**, 590–593.
33. Young TP, Chen BG, Kim TY *et al.* (2014) Finger counting: an alternative method for estimating pediatric weights. *Am J Emerg Med* **32**, 243–247.

34. Cattermole G, Leung MP, So H *et al.* (2010) Age-based formulae to estimate children's weight in the emergency department. *Emerg Med J* **28**, 590–593.

35. Lubitz DS, Seidel JS, Chameides L *et al.* (1988) A rapid method for estimating weight and resuscitation drug dosages from length in the pediatric age group. *Ann Emerg Med* **17**, 576–581.

36. Garland JS, Kishaba RG, Nelson DB *et al.* (1986) A rapid and accurate method of estimating body weight. *Am J Emerg Med* **4**, 390–393.

37. Wells M, Coovadia A, Kramer E *et al.* (2013) The PAWPER tape: a new concept tape-based device that increases the accuracy of weight estimation in children through the inclusion of a modifier based on body habitus. *Resuscitation* **84**, 227–232.

38. Cattermole G, Leung P, Mak P *et al.* (2010) Mid-arm circumference can be used to estimate children's weights. *Resuscitation* **81**, 1105–1110.

39. Whitfield KC, Wozniak R, Pradinuk M *et al.* (2016) Anthropometric measures are simple and accurate paediatric weight-prediction proxies in resource-poor settings with a high HIV prevalence. *Arch Dis Child* **102**, 10–16.

40. Wozniak R (2012) The evaluation of potential weight-estimation methods in a primarily HIV positive cohort in Botswana for use in resource limited settings. BSc Thesis, University of British Columbia.

41. Luscombe M & Owens B (2007) Weight estimation in resuscitation: is the current formula still valid? *Arch Dis Child* **92**, 412–415.