STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:2891–2911
Published online 18 August 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3667

Estimating regional centile curves from mixed data sources and countries

Stef van Buuren[1,2,*,†], Daniel J. Hayes[3], D. Mikis Stasinopoulos[4], Robert A. Rigby[4], Feiko O. ter Kuile[3] and Dianne J. Terlouw[3]

[1]*TNO Quality of Life, P.O. Box 2215, 2301 CE Leiden, The Netherlands*
[2]*Department of Methodology and Statistics, FSS, University of Utrecht, P.O. Box 80140, 3508 TC Utrecht, The Netherlands*
[3]*Child and Reproductive Health Group, Liverpool School of Tropical Medicine (LSTM), Pembroke Place, Liverpool L3 5QA, U.K.*
[4]*STORM Research Centre, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, U.K.*

SUMMARY

Regional or national growth distributions can provide vital information on the health status of populations. In most resource poor countries, however, the required anthropometric data from purpose-designed growth surveys are not readily available. We propose a practical method for estimating regional (multi-country) age-conditional weight distributions based on existing survey data from different countries. We developed a two-step method by which one is able to model data with widely different age ranges and sample sizes. The method produces references both at the country level and at the regional (multi-country) level. The first step models country-specific centile curves by Box–Cox t and Box–Cox power exponential distributions implemented in generalized additive model for location, scale and shape through a common model. Individual countries may vary in location and spread. The second step defines the regional reference from a finite mixture of the country distributions, weighted by population size. To demonstrate the method we fitted the weight-for-age distribution of 12 countries in South East Asia and the Western Pacific, based on 273 270 observations. We modeled both the raw body weight and the corresponding Z score, and obtained a good fit between the final models and the original data for both solutions. We briefly discuss an application of the generated regional references to obtain appropriate, region specific, age-based dosing regimens of drugs used in the tropics. The method is an affordable and efficient strategy to estimate regional growth distributions where the standard costly alternatives are not an option.

KEY WORDS: body weight; Asia; GAMLSS; centile estimation; mixture model; worm plot

*Correspondence to: Stef van Buuren, TNO Quality of Life, P.O. Box 2215, 2301 CE Leiden, The Netherlands.
†E-mail: stef.vanbuuren@tno.nl

Contract/grant sponsor: Drugs for Neglected Diseases initiative (DNDi); contract/grant number: LSTM2007
Contract/grant sponsor: Centers for Disease Control and Prevention (CDC); contract/grant number: 5U01/CI 000321
Contract/grant sponsor: Medical Research Council (MRC)/Liverpool School of Tropical Medicine (LSTM); contract/grant number: GO0501401

Received 26 June 2008
Accepted 11 June 2009

1. INTRODUCTION

A wide range of methods is available for estimating the age-related distribution of human growth [1, 2]. *Growth standards* are based on such distributions. Growth standards describe how children should grow, and are widely accepted tools for monitoring height, weight and body mass index. Many countries use standards developed for another country, such as the recently updated international WHO standards for children below 5 years of age [2] and standards for school-aged children and adolescents [3]. In contrast to growth standards, *growth references* describe the actual distribution of growth parameters at population level at a certain point or period in time, i.e. they show how children actually grow in a given population. Growth references provide vital information on the overall health status of a population and can be updated to show changes in health/growth over time [4].

In practice, growth references are rarely available in resource-poor regions since the required purpose-designed national or multi-country growth surveys are costly and logistically challenging. Some nutritional data may however be available from subgroups within the population from a variety of different sources. Large national household surveys such as UNICEF's multiple indicator cluster survey (MICS) program and the measure demographic health survey (DHS) program collect various nationally representative population-level representative demographic and health data. These studies currently typically cover anthropometric measurements in specific population subgroups only, such as children <3 or <5 years of age, or women of reproductive age (15–49 years). In addition, nutritional data that are representative at the sub-national/district level may be collected as a part of demographic surveillance systems or other large-scale research efforts.

Despite the availability of these data, there are currently no standardized methods that aim to capture multi-source data to derive reference curves. The use of existing data could be an affordable, efficient strategy when alternatives are not an option, but the success of this approach is highly dependent on the quality of the data used. Obvious problems include the incompleteness of data across age ranges, sex, ethnicity, geographic regions, time periods and/or socio-economic strata. In addition, differences in sampling design and sampling fraction, quality assurance of anthropometric measurements related to the accuracy of measurements, rounding or calibration of equipment would undermine this approach. With this incomplete list of potential drawbacks, one may be tempted to discard the use of existing data altogether. These problems, however, are not unlike those faced when conducting systematic reviews of the effects of health-care interventions. The Cochrane guidelines provide recommendations regarding literature searches, data inclusion, data extraction, quality assessment and the appropriate statistical methods of meta-analyses [5]. In the same way, when handled with care, the use of existing anthropometric data to estimate the distribution of human growth is likely to be informative and of public health relevance, despite their obvious limitations.

The methodology for creating growth references for individual populations for whom complete, quality data are available is well established [1, 6, 7]. The modeling of growth data from different sources and regions requires an expansion of current modeling methodology to address three main problems. First, populations may have different growth distributions with varying median, spread and skewness linked to genetic or nutritional factors. Second, existing data typically cover different age ranges and may not be equally available for both genders, whereas conventional methodology is limited to the available age range. The analysis may thus require data extrapolation. Third, country-specific references may need to be combined into a regional distribution covering multiple

DOI: 10.1002/sim

countries. This requires the overall distribution to be weighted for population size or other factors of interest.

In this paper we propose a practical method for estimating age-conditional distributions based on existing data from multiple sources. We will demonstrate the method by estimating the age-conditional distribution of body weight using existing data from countries in South East Asia and the Western Pacific. Section 2 introduces the data sources that were used to develop the method. Section 3 describes two different modeling options to estimate the age-conditional body weight distribution. Section 4 presents a method to combine individual country-specific distributions into a weighted regional distribution. Section 5 discusses some potential applications of the generated regional reference, in particular to calculate optimal age-based drug dosing regimens. Section 6 discusses the strengths and limitations of the method and concludes the paper. The appendix contains the R code for fitting the main models.

2. SOURCE DATA

Through literature searches and communication with individual researchers, research institutions and national and international agencies, we compiled community-based data from 12 Asian countries covering the WHO Western Pacific and South-East Asian regions. Body weight was measured for 273 270 individuals (62 523 males, 210 747 females) aged between 2 weeks and 50 years. Only population-based data sets collected as random samples of the population and direct weight measurements were included. Data based on self-reported body weight were excluded. Table I provides a breakdown of the number of observations per country. Detailed methods of the quality assessment, selection criteria and collation procedures of these different data sets will be described elsewhere.

All body weight measurements (in kg) were expressed as weight-for-age deviation scores, or Z scores, relative to the CDC 2000 growth reference [8]. The Z score for persons aged 20 years and older was calculated using the reference for the 20-years olds. Any weight gain in adults thus

Table I. Number of cases and age range per country and sex of South-East Asian countries.

		Females		Males	
Country	N	Age range (yrs)	N	Age range (yrs)	N
Bangladesh	17 486	0–4, 15–49	14 357	0–4	3129
Bhutan	368	36–49	212	36–49	156
Burma	8577	0–4	4278	0–4	4299
Cambodia	15 521	0–49	11 938	0–49	3583
India	165 576	0–49	141 307	0–49	24 269
Indonesia	2262	9–15	1127	9–15	1135
Laos	1547	0–4	787	0–4	760
Nepal	16 059	0–4, 15–49	13 362	0–4	2697
Philippines	730	7–30	279	7–31	451
Sri Lanka	2021	0–4	947	0–4	1074
Thailand	44 024	0–49	24 146	0–49	19 878
Viet Nam	8220	0–4, 9–10	3031	0–4, 9–10, 19–48	5189
Total	273 270		210 747		62 523

DOI: 10.1002/sim

implies a higher Z score. Outliers for weight were examined by first plotting the Z score by age for each study separately, followed by searching for discontinuities in the distributions. As the spread between countries was fairly large, we decided to apply liberal cutoff values of $Z<-10$ SDS and $Z>+6$ SDS for the automatic exclusion of outliers. Asymmetric cutoffs were chosen to reflect the fact that the mean weights of our data were considerably lower than the reference set (based on the U.S. population). The percentage of outliers thus removed was equal to 0.12 per cent of the data, and was no greater than 0.63 per cent for any one country.

The remaining observations were randomly split into two data sets: a *training* set of 70 per cent of the records for fitting and checking the model, and a *test* data set of 30 per cent of the records to assess the predictive validity and the risk of overfitting the obtained model.

Figures 1 and 2 contain simple scattergrams of body weight against age for the 12 countries. Note that each subplot exhibits the familiar pattern of increasing body weight with age, but there are substantive differences between countries with respect to the ages covered, and the density of the available data. In addition, some countries are covered by two or more studies. For example, one could recognize three different studies among Vietnamese males. In total 77 per cent of the data represented females. This unusual sex ratio is explained by the sampling strategy of the large scale, national DHS and MICS surveys, which contributed 83 per cent of the compiled data. In these surveys, anthropometric data are collected from children of both sexes up to the age of 3–5 years, from women of reproductive age, but not from adult males. This focus on adult females in these surveys is linked to the known association between maternal anthropometrics and child and reproductive health indicators.

3. MODELS FOR COUNTRY-LEVEL AGE-CONDITIONAL WEIGHT DISTRIBUTIONS

This section focuses on modeling the age-conditional body weight distribution for individual countries through a common model. The modeling framework is based on the generalized additive model for location, scale and shape (GAMLSS) [9]. GAMLSS provides a flexible model for a univariate response variable Y based on explanatory variables. The distribution of Y in GAMLSS can be selected from a large collections of distributions, including highly skewed and kurtotic distributions. The package for fitting GAMLSS is available in R [10]. The analysis presented here used the GAMLSS 1.8.7 library running under R 2.6.2.

Here, we consider a special case of the GAMLSS model, where we model the outcome Y using both a continuous explanatory variable x, a function of age and a categorical variable c for country (indexed by subscript j)

$$\begin{aligned} Y &\sim D(\mu, \sigma, \nu, \tau) \\ g_1(\mu) &= \alpha_{1j} + h_1(x) \\ g_2(\sigma) &= \alpha_{2j} + h_2(x) \\ g_3(\nu) &= \alpha_{3j} + h_3(x) \\ g_4(\tau) &= \alpha_{4j} + h_4(x) \end{aligned} \tag{1}$$

where D represents the distribution, either the Box–Cox t (BCT) distribution or the Box–Cox power exponential (BCPE) distribution. Equation (1) expresses the location (μ), scale (σ), skewness (ν) and kurtosis (τ) parameters of the distribution as smooth functions of explanatory variable x.

DOI: 10.1002/sim

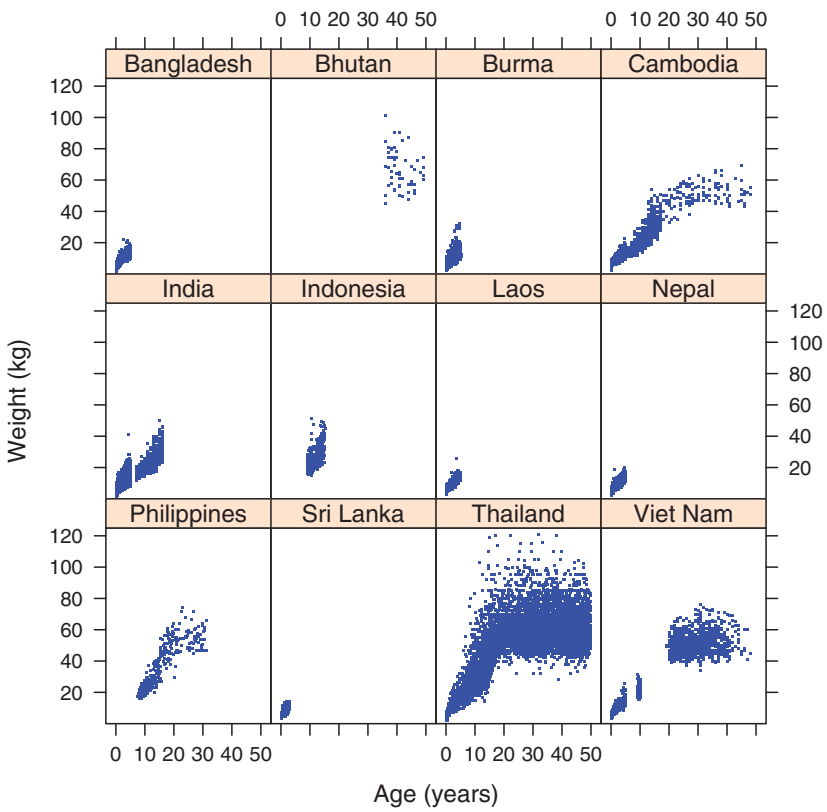

Figure 1. Weight (kg) by age (years) of males in 12 Asian countries. Sample sizes and age ranges for which data are available vary widely between countries.

The term α_{kj} is the constant intercept for country j for the distribution parameter $k(k=1,\ldots,4)$, $h_k(x)$ is either a polynomial or a smooth function of x and $g_k()$ is a known monotonic link function chosen by the user. Males and females are modeled separately.

We use the following stepwise procedure for selecting a country-level model:

1. Select Y (raw weight or weight-for-age Z score).
2. Select functions $g_k()$ and distribution D (BCPE or BCT).
3. Select transformation x of variable age.
4. Assess the need to split the countries into two or more homogeneous subgroups, if age-dependent weight distribution of subgroups is considerably different, using the Schwarz Bayesian information criterion (SBC).
5. For each homogeneous subgroup of countries, select size of penalty for each effective degree of freedom used in the models for μ, σ, ν and τ on the training data, for application with the generalized Akaike information criterion (GAIC), or a combination of worm plots and associated mean-squared error (MSE) plots.
6. Optimize models for each individual subgroup of countries.
7. Confirm overall model fit using the test data.

DOI: 10.1002/sim

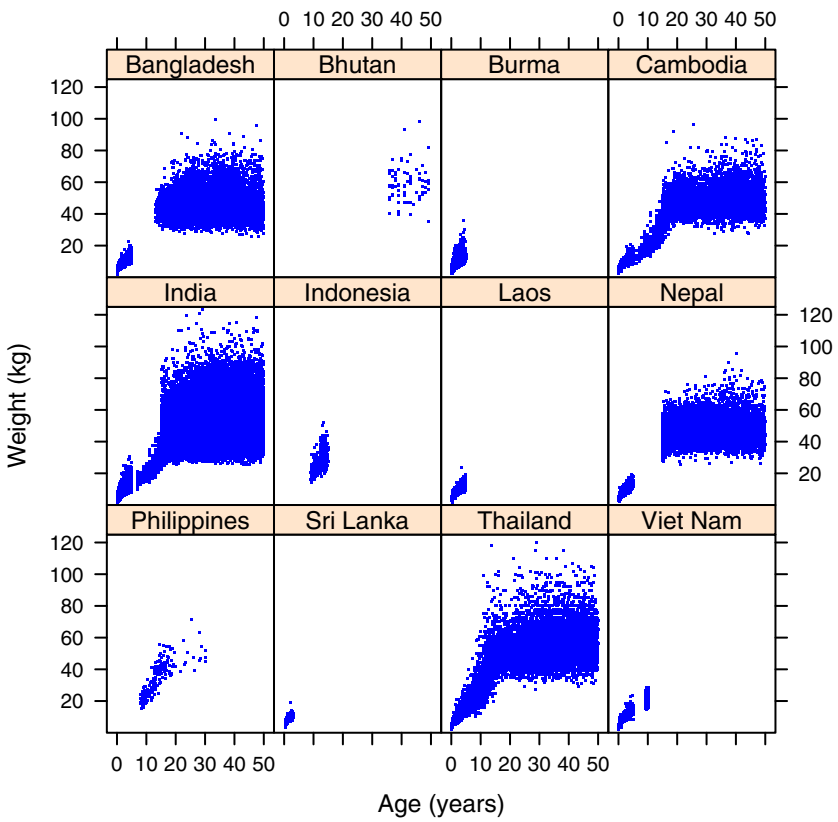

Figure 2. Weight (kg) by age (years) of females in 12 Asian countries. Sample sizes and age ranges for which data are available vary widely between countries.

Selecting Y, i.e. modeling W or Z: The function $Z = f(W,t)$ converts measurements in kg (W) into standard deviation scores (Z) relative to a chosen reference, while $W = f^1(Z,t)$ is its inverse. We present models for both W and Z. The first approach uses the raw weight in kg and simultaneously describes the variation in weight between different ages and between countries. The second approach models Z, and concentrates on modeling the variation between countries.

Distributions: Previous applications of GAMLSS include centile estimation using the BCPE and BCT distributions to a variety of anthropometric measures against age [11] and [12]. Both distributions are four parameter distributions, allowing modeling of location, scale, skewness and kurtosis. Both distributions allow modeling of positive and negative skewness. The difference between the BCPE and BCT distributions is that BCPE allows modeling of both platykurtosis and leptokurtosis, whereas BCT only models leptokurtosis. The BCT distribution has been found to provide a better fit to a response variable with high leptokurtosis, e.g. [12], whereas BCPE has the flexibility to fit a response variable with platykurtosis or moderate leptokurtosis.

Transformation of the age scale: To simplify the model needed to fit the median weight over the course of life, we applied a transformation to age t. In general, it is beneficial to expand the ages where growth velocity is high, and compress age where growth velocity is low [6]. Based on

DOI: 10.1002/sim

the optimized SBC, a simple square root transformation of age was found to appreciably enhance the model fit. We also tried the age transformation method proposed by Cole, whereby the median curve is first fitted against age in the usual way, then it is refitted but using the initial median curve rather than age. The refined median curve is then plotted against the original age [12]. However, this did not provide an improved fit to the data as judged by the SBC. All modeling was therefore done in the scale of $\sqrt{t}$.

Model selection: The optimal (effective) degrees of freedom in each of the smooth h functions above were selected based on the combination of two diagnostic tools:

(a) Minimization of the GAIC, given by $\text{GAIC} = -2l + \#\text{df}$, where l is the log likelihood function, # is a penalty for each degree of freedom used in the model and df is the total (effective) degrees of freedom used in the model. The penalty # chosen should be sufficiently large to avoid overfitting (which results in centile curves that are insufficiently smooth) and sufficiently small to avoid underfitting (which results in a poor fit to the data). We use the SBC, a specific GAIC where the penalty $\# = \log n$. As the penalty is related to the size of the data set, this reduces the risk of overfitting.

(b) In addition, worm plots per country and the associated MSE were used to guide these choices [7]. The worm plot is a de-trended QQ-plot created for a number of (usually 16) age intervals. The shapes of the worms indicate where and how we can improve the fit of the model. Worm plots were drawn on the level of the individual countries, and a customized software routine was created for this purpose. Our strategy involved the comparison of two worm plots side-by-side, where we gradually increased the degrees of freedom $\text{df}(\mu)$, $\text{df}(\sigma)$, $\text{df}(\nu)$ and $\text{df}(\tau)$. The degrees of freedom, df, were chosen at the point at which only very minor differences can be seen between the solutions df and df+1. In order to quantify this effect, we calculated the quantitative measures, β_0 to β_3, which describe the shape of each worm. The amount of fit of the μ-component in the model is measured by shape coefficient β_0. For age groups $g = 1, \ldots, 16$ we calculate the $\text{MSE}(\beta_0) = \Sigma_g \beta_{0,g}^2 / 16$ over the 16 panels of the worm plot. This is done separately for country and shape coefficients β_0 to β_3 on a grid of models for $\text{df}(\mu)$, $\text{df}(\sigma)$, $\text{df}(\nu)$ and $\text{df}(\tau)$.

The resulting local (country level) references were plotted in the form of age-dependent curves for μ, σ, ν and τ by using the `predict.gamlss` function.

In the following sections we present two alternative options to model either the body weight W against age (referred to as the 'raw score model'), or models the weight-for-age Z score against age (referred to as the 'Z score model').

3.1. Raw score model

Here, we model directly the body weights W against age and country using a GAMLSS model of type (1), applying the following steps:

1. Set distribution to BCT, and change the default link function from identity μ to $\log(\mu)$.
2. Set age transformation: $x = \sqrt{t}$.
3. Set age smoothing method: Cubic splines.
4. Assess the homogeneity between countries: Split the countries into two groups: Thailand versus the remaining countries (see below).
5. Optimize models for the two individual homogeneous subgroups using the GAIC (i.e. SBC), MSE and worm plots.

DOI: 10.1002/sim

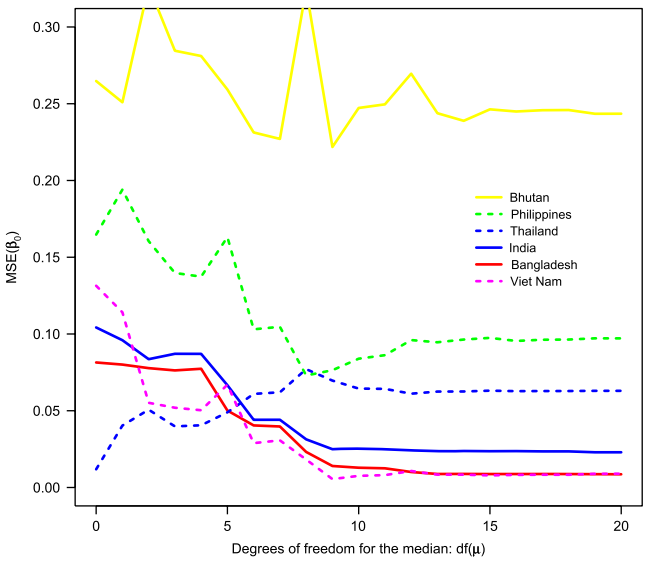

Figure 3. Mean-squared error (MSE) per country for increasingly flexible models for the median curve. MSE(β_0) quantifies model fit as the squared shape coefficient β_0 is averaged over 16 worm plot panels. The quantity df(μ) is the degrees of freedom in the cubic spline of $\sqrt{}$age in the median curve. A decreasing trend corresponds to a better fit. Note that the fit of Thailand deteriorates for more flexible models.

The rationale for this model choice is as follows. The BCT distribution was found to give a better overall fit to W than the BCPE distribution. The log link for μ was found to give a better fit than the identity link. The log link for μ provides a multiplicative model for μ, i.e. the countries have multiplicative effect on μ, while the identity link for μ would provide an additive model for μ. Cubic splines provided flexible smooth curves in age.

Figure 3 displays the MSE(β_0) per country for males at increasing degrees of freedom df(μ) before the split of the model into subgroups. For most countries the fit increases (i.e. the MSE decreases) as the model becomes more flexible. A notable exception is Thailand. We interpret this as an indication that the age-dependent weight distribution for males from Thailand is different from that in the other Asian countries. This is supported by the fact that the sum of the two deviances for the separate models is substantially lower than the deviance from the model fitted to all countries at once. We therefore decided to model Thailand separately. For the females, we came to a similar conclusion.

Hence, four models for W were fitted: male Thailand, male other countries, female Thailand and female other countries. All four models for W were in the form given by:

$$\begin{aligned} W &\sim \mathrm{BCT}(\mu, \sigma, \nu, \tau) \\ \log(\mu) &= \alpha_{1j} + h_1(x) \\ \log(\sigma) &= \alpha_{2j} + h_2(x) \\ \nu &= h_3(x) \\ \log(\tau) &= h_4(x) \end{aligned} \tag{2}$$

DOI: 10.1002/sim

Table II. Degrees of freedom of smooth functions of $x=\sqrt{\text{age}}$ that describes the age-varying distribution of body weight (kg) as the four components in the BCT model.

	df(μ)	df(σ)	df(ν)	df(τ)
Male Thailand	14	7	4	0
Male other countries*	13	7	3	0
Female Thailand	12	6	2	0
Female other countries*	20	12	10	3

*Other countries' model include data from Bangladesh, Bhutan, Burma, Cambodia, India, Indonesia, Laos, Nepal, Philippines, Sri Lanka and Viet Nam. Cubic splines were used. The default link function was changed from identity μ to $\log(\mu)$.

The male, 'other countries', model includes different constant intercepts α_{1j} and α_{2j} for each of the other countries (indicated by subscript j) for each of μ and σ, respectively. Similarly for the female, 'other countries', model. The male and female Thailand models are in the same form, but without the country effect, i.e. setting $\alpha_{1j}=0$ and $\alpha_{2j}=0$.

The functions $h_k(x)$ for $k=1,2,3$ are smooth functions of x for all four models, while function $h_4(x)$ is linear in x, i.e. $h_4(x)=\alpha_4+\beta_4 x$ for all models except the female 'other countries' model where it is a smooth function of x.

Table II provides the chosen degrees of freedom for smoothing (on top of constant and linear terms in x) for the h functions of μ, σ, ν and τ. Note that all four models have different fitted smooth functions and the fitted linear models for $h_4(x)$ in three of the models (where the smoothing degrees of freedom are zero) are also different.

The use of the log link for μ affects the interpretation of the median curve. For the other countries model (2) we have

$$\mu = e^{\alpha_{1j}} e^{h_1(x)} = \mu_j H_1(x)$$

$$\sigma = e^{\alpha_{2j}} e^{h_2(x)} = \sigma_j H_2(x)$$

where $H_1(x)=e^{h_1(x)}$ and $H_2(x)=e^{h_2(x)}$ are the smooth functions of x, while $\mu_j=e^{\alpha_{1j}}$ and $\sigma_j=e^{\alpha_{2j}}$ are the scaling factors for μ and σ for the country j ($j=1,\ldots,J$). Hence, the fitted functions for μ (the median weight) against age (for the different other countries) differ only in their scaling factor μ_j.

Figure 4 plots the estimates of μ_j against σ_j. The medians for Bhutan and Philippines were about 1.1–1.3 times the common median curve. The data for India had a high variance, whereas the spread in the data was low in Viet Nam.

3.2. *Z score model*

This section discusses modeling Z instead of W. The Z measure expresses all data relative to the CDC 2000 weight-for-age references. Fitting the model involved the following steps:

1. Set distribution to BCPE, using default link functions.
2. Set age transformation: $x=\sqrt{t}$.
3. Set age smoothing method: orthogonal polynomials for model exploration, switch to cubic splines for final models.

DOI: 10.1002/sim

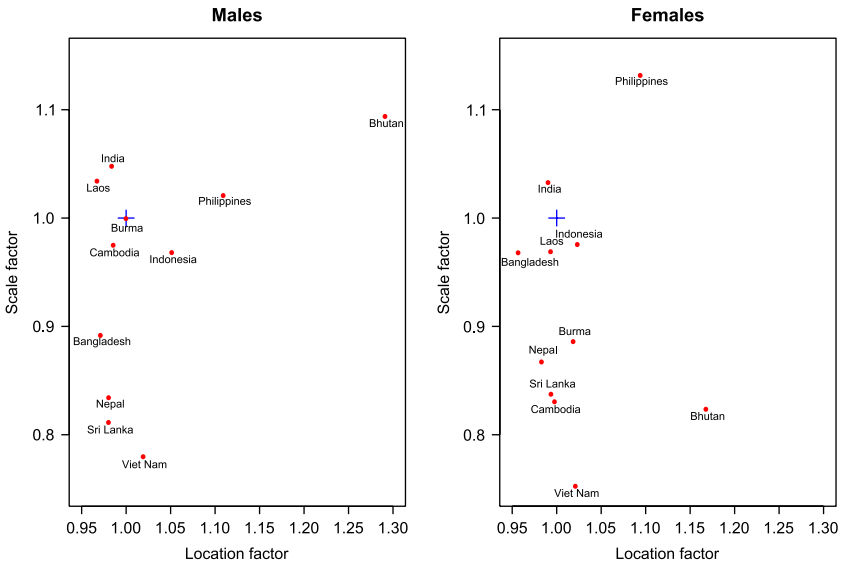

Figure 4. Scattergrams of the two country parameters (location and spread) for the raw score model, by sex. The horizontal axis is a multiplicative location factor of the common μ-component. A larger value indicates a higher median weight of the data. The vertical axis is a multiplicative scale factor of the common σ-component. A larger value implies a higher level of variability in weight in the data.

4. Add a constant = 11 to the Z score data before fitting (see below for rationale).
5. Split the countries into two groups: Thailand versus the rest.
6. Optimize models using the GAIC (i.e. SBC), MSE and worm plots.

The distribution of Z is platykurtic, i.e. it has thicker tails than the normal distribution. The BCPE distribution is able to model this feature of the data [11]. As the BCPE is only defined for positive values, we add a constant $\delta = 11$ to the data before fitting, based on the minimum deviance found over a grid of δ-values using the profile likelihood.

All models for Z are given by

$$\begin{aligned} Z &\sim \text{BCPE}(\mu, \sigma, \nu, \tau) \\ \mu &= \alpha_{1j} + h_1(x) \\ \log(\sigma) &= \alpha_{2j} + h_2(x) \\ \nu &= h_3(x) \\ \log(\tau) &= h_4(x) \end{aligned} \tag{3}$$

with $h_1(x)$ and $h_2(x)$ are the smooth functions of x whose degree of freedom is to be specified, and where $h_3(x) = \alpha_3 + \beta_3 x$ and $h_4(x) = \alpha_4 + \beta_4 x$ are linear in x. As results were largely insensitive to the type of smoother, we chose the fastest method (in terms of computer processing time), orthogonal polynomials, for model exploration, and cubic splines for the final models. The model allows for different constant intercepts α_{1j} and α_{2j} for the countries in the models for each of

DOI: 10.1002/sim

Table III. Degrees of freedom of smooth functions of $x = \sqrt{\text{age}}$ that describes the age-varying distribution of the Z score of body weight as the four components in the BCPE model.

	df(μ)	df(σ)	df(ν)	df(τ)
Male Thailand	8	1	*	1
Male other countries†	8	2	*	1
Female Thailand	5	2	1	1
Female other countries†	12	2	*	1

*includes only the intercept α_3.

†Other countries' model include data from Bangladesh, Bhutan, Burma, Cambodia, India, Indonesia, Laos, Nepal, Philippines, Sri Lanka and Viet Nam. Cubic splines were used. Z scores were calculated relative to the CDC 2000 reference. The BCPE model was fitted to the $Z+11$.

μ and σ, respectively. The model allows that countries can differ in their mean level and their variation on the Z score scale, but not in other aspects.

The model requires a specification of the degrees of freedom, df(μ), df(σ), df(ν) and df(τ). It is essential to assess the fit of the models at the country level since the composition of the sample can differ from that of the regional population. We studied the behavior of the worm plots at the country level, with its vertical range set equal to -0.5–$+0.5$. Smoothing coefficients were fixed at the point at which no major improvements in fit occurred anymore for successively df(μ), df(σ), df(ν) and df(τ). Table III lists the settings of the Z score models. After fitting, all results were back transformed into the original body weight scale.

Figure 5 is a scattergram of the estimates of μ_j against σ_j. For males, the median weight of the country is between -1 SD and -2 SD below the CDC reference. For females, the range drops between -2 SD and -3 SD, with Bangladesh females being the lightest of all at -2.78 SD. The variance Z is high for Laos and India, and low for Viet Nam.

3.3. Comparison of the raw score and Z score models to define Y

Figure 6 shows the fitted median weight curves of six countries under both models. In the raw score models, the curves are proportional to each other, except for Thailand. In the Z score model, the medians are shifted versions of each other in the Z scale, again except for Thailand. Despite substantial differences in the model, both methods resulted in very similar estimates of the age-conditional weight distribution. For external comparison, Figure 6 includes the median and the -2 SD line of the CDC references.

We also compared the worm plots and the centiles of the raw score and Z score models. As an example, Figure 7 contains the worm plots for Bangladesh males under both models. The use and interpretation of worm plots as diagnostic tools for modeling growth reference curves have been described in detail elsewhere [8]. While some differences are certainly present, both model solutions are largely, if not very comparable.

Table IV contains the percentage of persons that fall below the 3rd, 10th, 50th, 90th and 97th centiles of the country for both models. We calculated the percentages from the test data, which were not used to estimate the model. Table IV indicates that most percentages are located within 1 or 2 per cent points of the expected value. Larger differences may occur when sample size is low (e.g. Philippines). It was not possible to estimate the percentages for Bhutan. There is not much difference between the raw score and Z score models. Table IV provides evidence that the fitted country-level references adequately model the empirical distributions in each country.

DOI: 10.1002/sim

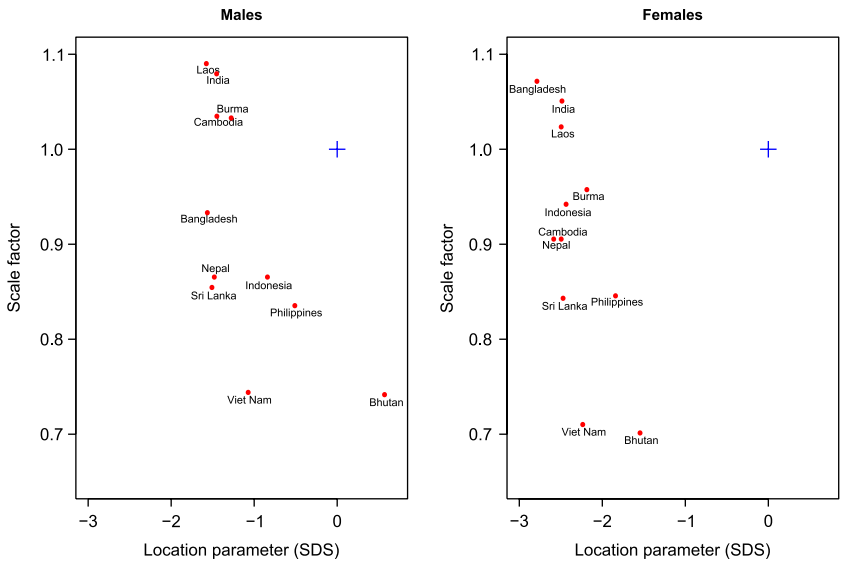

Figure 5. Scattergrams of the two country parameters (location and spread) for the Z score model, for males and females. The horizontal axis is the shift in location (in SDS) relative to the CDC 2000 reference. A larger value indicates a higher median weight of the data. The vertical axis is a multiplicative scale factor of the common σ-component. A larger value implies a higher level of variability in weight in the data.

4. COMBINING COUNTRIES INTO A REGIONAL WEIGHT-FOR-AGE DISTRIBUTION

On completion of the country-level modeling, the regional distribution is determined by mixing the smoothed country-level distributions. To do this, let $f_j(y)[=f_j(y|X=x)]$ be the distribution of Y given $X=x$ for country j, for $j=1,2,\ldots,J$. The regional distribution $f_Y(y)[=f_Y(y|X=x)]$ is a finite mixture distribution with known mixture weights w_j proportional to the population size P_j, given by

$$f_Y(y)=\sum_{j=1}^{J} w_j f_j(y) \quad \text{where } w_j=\frac{P_j}{\sum_{j=1}^{J} P_j} \text{ for } j=1,2,\ldots,J \tag{4}$$

Similarly, the regional cumulative distribution function $F_Y(y)[=F_Y(y|X=x)=\text{prob}(Y\leqslant y|X=x)]$ is given by

$$F_Y(y)=\sum_{j=1}^{J} w_j F_j(y) \tag{5}$$

where $F_j(y)$ is the cumulative distribution function of Y given $X=x$ for country j. The regional centile y given x, for a specific centile percentage $100p$ per cent, is defined by $F_Y(y)=p$, or equivalently $y=F_Y^{-1}(p)$, requiring the inverse cumulative distribution function $F_Y^{-1}(p)$. A centile curve for y against x, for a specific centile percentage $100p$ per cent, is obtained by finding

DOI: 10.1002/sim

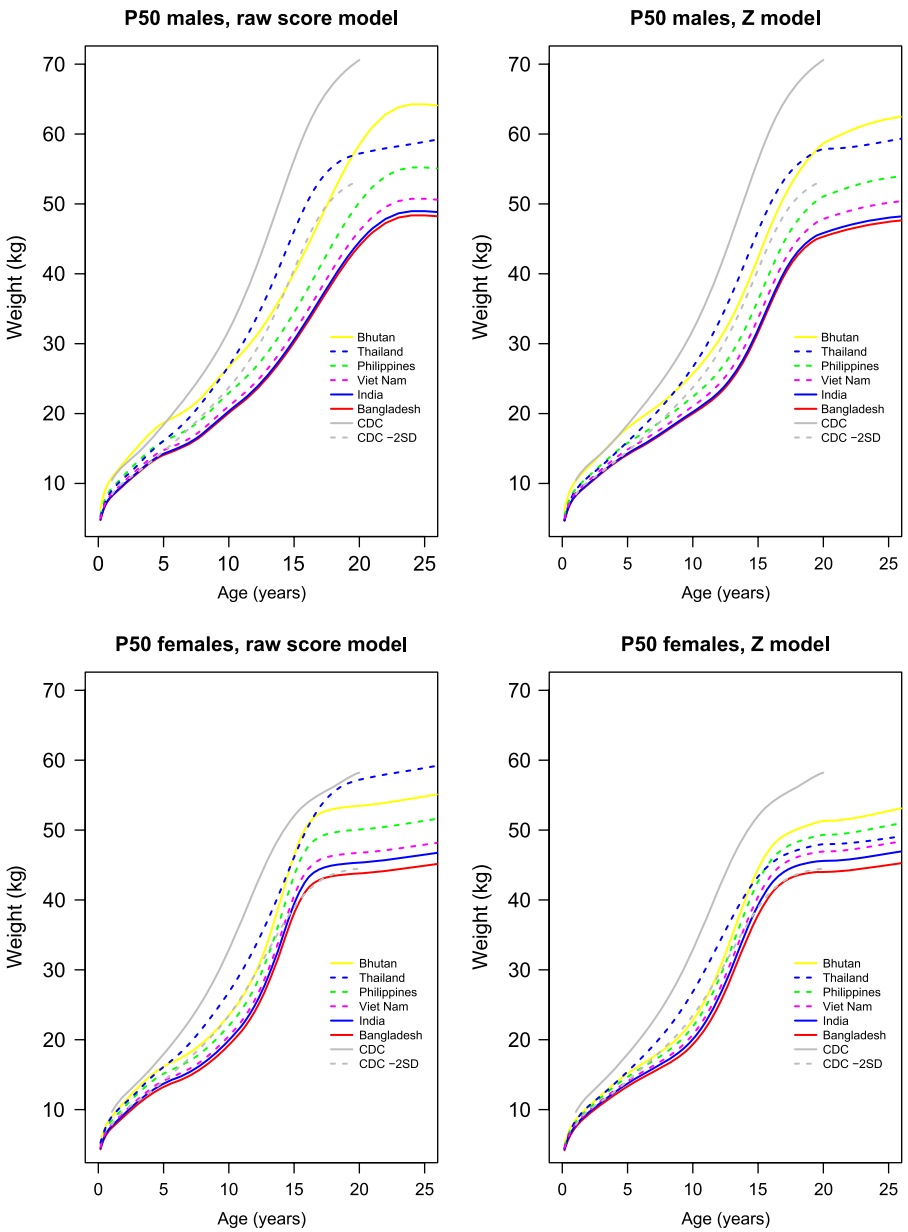

Figure 6. Median ($P50$) weight-by-age curves per country by model types (raw score and Z score) and sex. For comparison purposes, the median and −2SD lines of the CDC 2000 reference have been added. Note that the Asian people are on average substantially lighter than the CDC 2000 reference.

the centile y for a sequence of values of x. In the raw score model, we mix the country-level distribution of W. In the Z score model, the mixture is created using the weight-for-age Z score, and subsequently back transformed into the kg scale.

DOI: 10.1002/sim

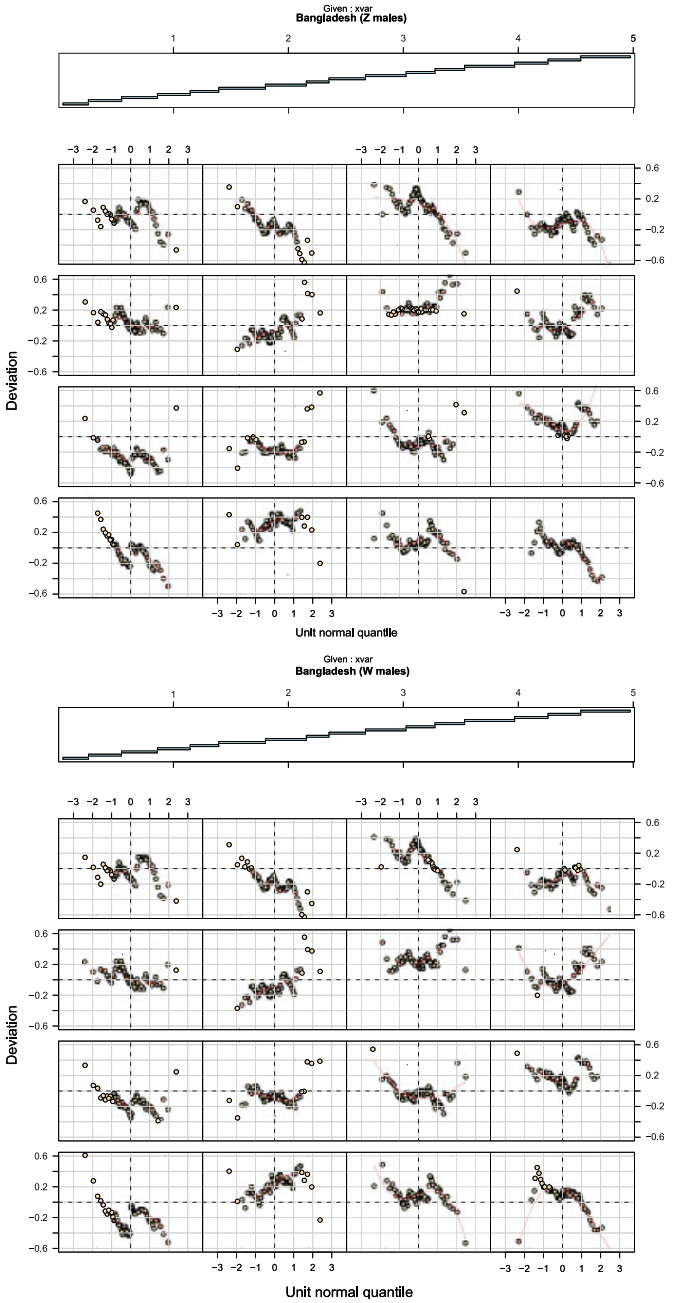

Figure 7. Example of worm plot used to assess the accuracy of the model fit for males in Bangladesh based on the test data. 'Flatter worms' indicate a better fit. The fit of the Z score model (top) and the raw score model (bottom) are similar. Similar age-stratified diagnostic plots were used by country and sex. Age groups are ordered from the lower-left corner panel to the upper-right corner panel.

DOI: 10.1002/sim

Table IV. Percentages of cases below the $P3$, $P10$, $P50$, $P90$ and $P97$ centiles per country in the test data for the raw score model and the Z score model. The percentages for Bhutan could not be estimated due to the low number of cases.

	Raw score (W) model					Z score model				
Country	$P3$	$P10$	$P50$	$P90$	$P97$	$P3$	$P10$	$P50$	$P90$	$P97$
Males										
Bangladesh	2.7	9.8	50.9	90.2	96.8	2.6	9.3	51.1	90.2	96.9
Burma	3.2	9.9	48.2	91.0	96.8	2.7	9.6	48.6	92.2	97.1
Cambodia	3.5	9.4	49.5	90.8	97.5	3.1	10.5	50.8	91.0	97.5
India	3.1	10.5	50.6	90.6	97.3	3.0	10.1	50.3	90.9	97.5
Indonesia	1.3	8.9	50.5	90.8	96.8	1.3	9.8	50.5	92.4	97.1
Laos	3.2	11.3	52.3	89.2	95.0	2.7	10.8	51.8	91.0	95.5
Nepal	3.5	10.0	48.6	89.6	97.0	3.8	10.2	48.4	89.9	97.5
Philippines	4.3	11.4	45.7	89.3	99.3	4.3	10.7	45.7	90.0	99.3
Sri Lanka	3.3	9.9	45.9	91.3	99.4	3.0	9.3	44.7	91.6	98.8
Thailand	2.8	9.3	51.6	89.6	97.1	2.3	9.2	51.9	89.7	97.0
Viet Nam	2.5	9.5	49.0	90.7	97.6	2.6	8.8	49.3	90.8	97.6
Females										
Bangladesh	2.7	10.4	51.7	89.6	96.8	2.4	9.6	51.4	89.4	96.7
Burma	3.5	10.3	48.6	91.4	97.8	2.3	8.5	50.5	91.9	97.3
Cambodia	2.8	10.2	51.5	90.6	97.5	3.3	9.3	49.1	90.9	97.7
India	2.7	9.8	49.9	89.9	97.2	2.7	9.6	49.8	90.0	97.1
Indonesia	1.4	11.1	49.7	88.1	97.0	1.1	11.9	49.7	88.9	95.8
Laos	2.5	7.5	51.7	90.0	96.0	2.5	8.5	50.2	91.0	96.5
Nepal	3.1	9.8	49.9	89.7	97.0	3.0	9.8	49.4	90.1	97.4
Philippines	6.0	10.7	57.1	94.0	96.4	3.6	15.5	57.1	94.0	97.6
Sri Lanka	2.6	10.7	50.2	90.6	97.4	2.6	10.5	51.3	90.3	96.6
Thailand	3.2	10.0	51.4	90.2	96.9	3.1	9.6	51.5	90.2	97.0
Viet Nam	2.9	8.9	48.3	88.8	96.8	2.5	8.2	49.4	88.4	96.0

The generated regional centile curves are given in Figure 8 for ages from birth to 50 years. All in all, the regional reference from the raw score and Z score model agrees quite close, except for the 18-year old males. The primary reason for this is that the available male data are very sparse at this age (cf. Figure 1). Thus, the local shape of the model is not steered by empirical evidence. Our use of two distinct models highlights where in the model the solution seems to depend on assumptions. Provided that the models are flexible enough, we expect such discrepancies to disappear for more dense data, e.g. as was indeed the case in the females.

5. APPLICATION

The generated growth references may have a range of practical applications. Examples include monitoring the nutritional status in a country or a region, estimating age-specific prevalence of underweight or overweight and setting regulations that involve body weight, e.g. maximum number of persons allowed in a transport vehicle or elevator.

We developed the method with the specific aim to establish a standardized tool to determine safe and effective age-based dosing regimens for antimalarials and other drugs used in the tropics

DOI: 10.1002/sim

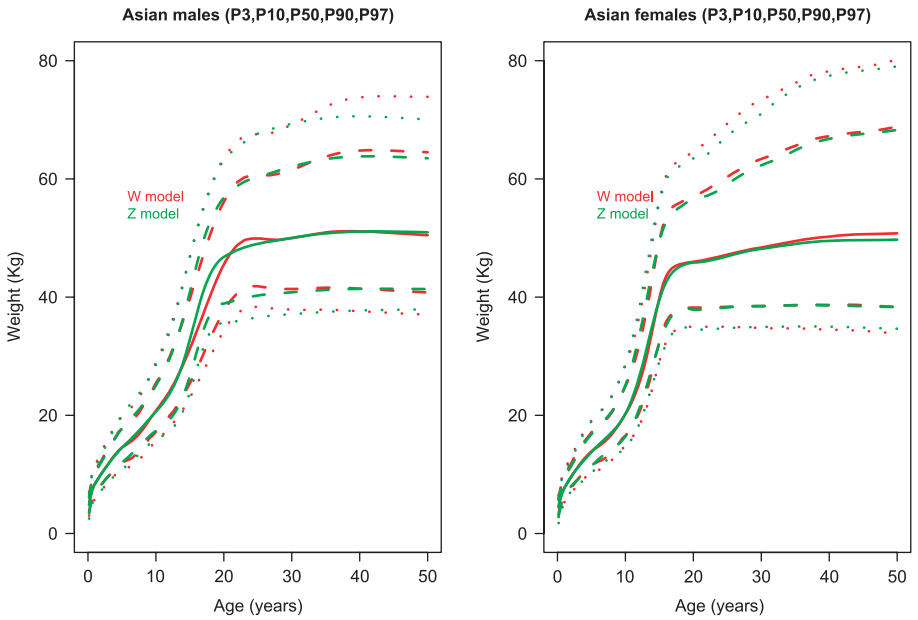

Figure 8. Regional centiles ($P3$, $P10$, $P50$, $P90$, $P97$) obtained after weighting the fitted country-specific distributions according to population size. Centiles from the raw score (W) model (dark red) and Z score model (light green) are plotted in the same figure. Males (left) and females (right).

[13]. Malaria causes an estimated 300–500 million episodes of clinical malaria, resulting in up to 1 million deaths each year. The regulatory drug development process for antimalarials typically results in weight-based dosing recommendations. In practice, however, the majority of fevers in malaria endemic areas are treated with over-the-counter antimalarial drugs without involvement of the formal health sector, or in other situations where functioning weighing scales are not available [14]. In these settings age is widely used as a proxy for weight. There are currently no standardized procedures to devise age-based proxies for the weight-based recommendations that typically result from the regulatory drug development process. This has contributed to considerable variability in existing age-based dose regimens for antimalarials, at times resulting in poor, but widely used regimens, particularly in those who bear the brunt of the malaria burden, namely young, rapidly growing children [15]. An optimal age-based regimen corresponds to the lowest risk of over- and under dosing in a given population. In order to derive optimal regimens, we need the actual reference distribution of weight by-age in that population, rather than the available international growth standards. The proposed method provides a tool to estimate these age-conditional distributions of body weight based on the available data.

6. DISCUSSION

We developed a two-step method for fitting centile growth curves for a set of countries. Our method is able to model data with widely different age ranges and sample sizes, and produces references both at the country level and at the regional level. Within the model, individual countries can vary

DOI: 10.1002/sim

in location and spread. We demonstrated the method by fitting the weight-for-age distribution in 12 countries in South East Asia and the Western Pacific, and were able to achieve a good fit between the final model and the original data.

Our two-step finite mixture approach is an improvement over methods that require pooled data for a single population. The first step generates country-specific references that extrapolate information from similar countries. Given sufficient overlap in the ages, this step allows for extrapolation of the country references into age ranges that were not observed. The second step defines the regional reference as a finite mixture distribution of the country references. We used mixture weights proportional to country-level population size, but other choices are certainly possible (e.g. malaria prevalence).

The two-step approach allows for tremendous flexibility and could be useful for mixing growth references within other settings. For example, suppose that we want to create a smooth transition between two disjoint references on a given age range. We could chose mixture weights to gradually change with age, thus setting $(1, 0)$ at the start of the age range setting $(0, 1)$ at the end, with a smooth transition in between. Alternatively, we could mix different subsamples (e.g. ethnic background, breast feeding status) in the appropriate proportions, or estimate secular trend in the distribution as one or more model parameters.

The country references are well-defined BCPE or BCT distributions, but the regional mixture distribution is not. Whether this is a problem depends on the context. In our application for defining age-based drug regimens, it was straightforward to calculate the percentage of adequately dosed children from the regional centiles. Calculating Z scores relative to the new regional reference would require a BCT or BCPE distribution. We fitted BCPE and BCT models to the centiles that approximate the regional reference to a high degree of accuracy. This fitted distribution could then be used to calculate Z scores.

We used two definitions for Y, the raw score and the Z score. Despite some methodological differences, both approaches resulted in very similar regional and country-level estimates, adding credibility to the validity and robustness of the procedure. Both approaches appear to have sufficient flexibility built into the model, which means that the results are predominantly dictated by the underlying data. Of course, this also implies that, as always, care is needed when data are sparse. The choice between modeling raw scores or Z scores is primarily a matter of convenience and familiarity. Modeling the raw score provides a model in terms of original units, which is convenient. Furthermore, it is the only option if appropriate external references are lacking. The Z score model scales data to an external growth reference, so that the location of a country can be expressed relative to this reference. The Z score model does not need to model the typical age-related shapes (e.g. the growth spurts during infancy and puberty, adiposity rebound) that appear in the original scale, thus allowing the use of a simpler model. On the other hand, the Z score approach may have difficulty in modeling populations with widely different timings. A prerequisite is that the reference used to obtain the Z scores is in phase with the new data, so that for example the timing of the pubertal growth spurt is similar of the two. If pubertal growth occurs earlier in the target population than in the reference population, this would manifest itself as a rise in Z scores around the start of puberty in the target population, followed by a decrease and leading up to an asymptote that could differ from the pre-pubertal mean. If the timing differs between different countries, then smoothing the Z scores tends to flatten out phase differences. This might have been the case for the males in Figure 6, where the curves around age 15 are different in shape under the two models. We cannot be sure, however, since there are only few data points around that age. In principle, we can allow for phase variation by using more flexible smoothing methods that incorporate time

DOI: 10.1002/sim

shifts, but that requires larger samples around puberty than are currently available. It is not known whether the raw score model is more efficient in handling this problem than the Z score model. A third option is to follow both routes, as presented here. This is more work, but it provides insight into the robustness of the final model under alternative modeling choices.

The age range for the analysis is 0–50 years. Some countries (Cambodia, Thailand, Bangladesh, India and Nepal) have substantial numbers of adult subjects. One concern is that those subjects may dominate the analysis, and swamp the child data. We reran our final models on the data that exclude any persons >25 years, and compared the results with our original curves. Overall, the results were very similar with only some small differences near the ages close to 25 years, presumably due to edge effects. We prefer the analysis 0–50 years since that analysis uses all available data.

The Z score method requires a growth reference. We have chosen to use the CDC 2000 reference, which is a country-specific reference based on the 1977 NCHS data. We initially planned to use the new WHO 2007 re-analysis of the NCHS data [3]. The WHO 2007 weight-for-age references only includes children up to the age of 10 years, however, and fails to cover the age interval 10–19 years. We therefore opted to use the CDC reference instead. The choice of reference, however, does not appear to be critical. Repeating the analysis using the Dutch 1997 reference instead of the CDC reference, [16] resulted in different solutions when viewed in the Z scale (as expected), but similar when those results were transformed into the W scale. It thus appears that the Z score method is robust against the choice of the reference.

As part of the model building process we explored the need to fit more complex models. Varying slopes for age were initially included in the raw score model for μ in the female data for Bangladesh, Cambodia and Nepal. Though this significantly improved model fit, the effect was only noticeable for the upper age range. The absolute difference of the fitted median was 0.28 kg when averaged over all ages, and 0.1 kg when averaged for ages 0–20 years. The largest deviation, 1.6 kg, occurred for Bangladesh females aged 50 years. As the differences of this magnitude are unlikely to have practical consequences, we refrained from modeling varying slopes in the final model. Likewise, other model parameters that we explored, such as study period and country effects for the models of ν and τ, had only minor effects and were therefore omitted.

Our method uses readily available data, which is an affordable and efficient strategy. We are, however, cognizant of the limitations of the generated growth reference curves. As with any model, the quality of the modeled output depends on the quality and representativeness of the underlying data. Though we included only studies where sampling was based on random sampling, the country-level curves we derived may not be representative if data are missing over a large age range for that country and if the growth rate in the missing age group differs from that in other countries contributing to the model. Furthermore, due to the sparsity of data from certain countries, we did not restrict ourselves to recent studies only. While most studies were conducted in the last 10 years, there were a few older surveys. The generated reference curves thus need to be updated as more recent, complete data become available.

Asia is home to a vast and diverse portion of the world's population, with differences in weight distribution within- and between-Asian countries linked to ethnicity, social and economic conditions, differences in diet, degrees of urbanization and nutritional transitions in recent times. These factors may have contributed to the observed differences in weight distribution of the Thai data versus the other malaria endemic countries from the region that was included in our analyses. Our regional reference curves also quantify known, but nevertheless striking differences of the

DOI: 10.1002/sim

actual Asian growth relative to the CDC 2000 growth references. Over 50 per cent of the population in Asia is estimated to have weights that are below the 3rd percentile of the CDC references.

There are initiatives to conduct national household surveys like UNICEF's MICs surveys and Measure DHS surveys on a more regular basis (every 3–4 years). Regular high-quality national household surveys would offer the opportunity to expand their current collection of representative anthropometric data in young children and women of reproductive age, to a wider age range. If such data were become available in the near future, they could readily be taken up into our approach. Doing so would help strengthen and solidify the information base of regional growth estimates, and potentially provide a tool to monitor country-level and regional changes in nutritional status. Until then, our method based on existing anthropometric data is an affordable and efficient strategy to monitor nutrition and health at several levels.

APPENDIX A: R CODE FOR FITTING MODELS

Example of data structure of Asian males:

```
ID  country   age weight   WAZ    WAZ11

3  Viet Nam   4.79 50     -2.514   8.486

9  Viet Nam   5.29 52      -2.186   8.814

10 Viet Nam   5.56 69     -0.144 10.856
```

where the column `age` indicates the square root of age in years, where WAZ is the Z score and where WAZ11 = WAZ + 11. The raw score model and the Z score model for the 'other countries' (excluding Thailand) were fitted in R by:

```
library(gamlss)

fit.w <- gamlss(formula = weight ~ cs(age, df = 13) + country,

          sigma.formula = ~ cs(age, df = 7) + country,

          nu.formula = ~ cs(age, df = 3, c.spar = c(-1.5,2.5)),

          tau.formula = ~ cs(age, df = 0),

          family = BCT(mu.link = "log"),

          data = asianMRest, c.crit = 0.1)

fit.z <- gamlss(formula = WAZ11 ~ cs(age, df = 8) + country,

          sigma.formula = ~ cs(age, df = 2) + country,

          nu.formula = ~ 1,
```


DOI: 10.1002/sim

```
tau.formula = ~ cs(age, df = 1),

family = BCPE,

data = asianMRest, c.crit = 0.1)
```

ACKNOWLEDGEMENTS

Part of the data used was available in the public domain. Other databases were kindly provided by different investigators. We are very grateful to the following scientists and institutes for sharing their data for the development of this methodology: Dr Sureeporn Punpuing: Kanchanaburi DSS, Institute for Population and Social Research, Mahidol University at Salaya, Nakhonpathom, Thailand; Measure DHS, MACRO International, Calverton, U.S.A.; UNICEF, New York, U.S.A.; Human Systems Integration Information and Analysis Centre (HSIIAC), Ohio, U.S.A.; Dr Simon Brooker: Department of Infectious and Tropical Disease, London School of Hygiene and Tropical Medicine (LSHTM), London, U.K.; Dr Tjalling Leenstra, Academic Medical Centre, Amsterdam, The Netherlands; and Dr Antonio Montresor, Western Pacific Regional Office of WHO (WPRO), Manila, Philippines. We acknowledge Piero Olliaro (UNICEF/UNDP/World Bank/WHO Special Programme on Research & Training in Tropical Diseases TDR) for our ongoing collaborative efforts on developing standardized methods for age-based dosing regimens. We thank the reviewers for their insightful comments.

This work was financially supported by the Drugs for Neglected Diseases initiative (DNDi) (grant number 'LSTM2007'). Dianne Terlouw and Feiko ter Kuile acknowledge further financial support from the Centers for Disease Control and Prevention (CDC) (Grant number 5U01/CI 000321). Daniel Hayes acknowledges support from a joint Medical Research Council (MRC)/Liverpool School of Tropical Medicine (LSTM) PhD fellowship (Grant number GO0501401).

REFERENCES

1. Wright EM, Royston P. A comparison of statistical methods for age-related reference intervals. *Journal of the Royal Statistical Society*, *Series A* 1997; **160**:47–69.
2. Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L, van Buuren S, Pan H, Molinari L, Martorell R, Onyango A, Martines J. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Statistics in Medicine* 2006; **25**:247–265.
3. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization* 2007; **85**:660–667.
4. Tanner JM. Use and abuse of growth standards. In *Human Growth*: *A Comprehensive Treatise. Volume 3*: *Methodology and Ecological*, *Genetic*, *and Nutritional Effects on Growth* (2nd edn), Falkner F, Tanner JM (eds). Plenum Press: New York, 1986.
5. Higgins JPT, Green S. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0*. [updated February 2008], The Cochrane Collaboration, 2008.
6. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine* 1992; **11**:1305–1319.
7. van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine* 2001; **20**:1259–1277.
8. Ogden CL, Kuczmarski RJ, Flegal KM, Mei Z, Guo S, Wei R, Grummer-Strawn LM, Curtin LR, Roche AF, Johnson CL. Centers for disease control and prevention 2000 growth charts for the United States: improvements to the 1977 national center for health statistics version. *Pediatrics* 2002; **109**:45–60.
9. Rigby RA, Stasinopoulos DM. Generalized additive models for location scale and shape (with Discussion). *Applied Statistics* 2005; **54**:507–554.
10. Rigby RA, Stasinopoulos DM. Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software* 2007; **23**:7.
11. Rigby RA, Stasinopoulos DM. Using the Box–Cox t distribution in GAMLSS to model skewness and kurtosis. *Statistical Modelling* 2006; **6**:220–229.

DOI: 10.1002/sim

12. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine* 1998; **17**:407–429.
13. Taylor WR, Terlouw DJ, Olliaro PL, White NJ, Brasseur P, ter Kuile FO. Use of weight-for-age-data to optimize tablet strength and dosing regimens for a new fixed-dose artesunate–amodiaquine combination for treating Falciparum malaria. *Bulletin of the World Health Organization* 2006; **84**:956–964.
14. Foster S. Treatment of malaria outside the formal health services. *Journal of Tropical Medicine and Hygiene* 1995; **98**:29–34.
15. Terlouw DJ, Courval JM, Kolczak MS, Rosenberg OS, Oloo AJ, Kager PA, Lal AA, Nahlen BL, ter Kuile FO. Treatment history and treatment dose are important determinants of sulfadoxine–pyrimethamine efficacy in children with uncomplicated malaria in Western Kenya. *Journal of Infectious Diseases* 2003; **187**:467–476.
16. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatric Research* 2000; **47**:316–323.

DOI: 10.1002/sim