

A Bayesian approach to estimate the prevalence of low height-for-age from the prevalence of low weight-for-age

Uma abordagem Bayesiana para obter estimativas de prevalência de déficit de crescimento com base na prevalência de baixo peso para idade

Michael E. Reichenheim ¹
Nicola G. Best ²

¹ Departamento de Epidemiologia, Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro. Rua São Francisco Xavier 524, 7ª andar, Rio de Janeiro, RJ, 20559-900, Brasil. michael@ims.uerj.br
² Department of Epidemiology and Public Health, Imperial College School of Medicine. St. Mary's Campus, Norfolk Place, London, W2 1PG, UK. n.best@ic.ac.uk

Abstract Victora et al. (1998) proposed the use of low weight-for-age prevalence to estimate the prevalence of height-for-age deficit in Brazilian children. This procedure was justified by the need to simplify methods used in the context of community health programs. From the same perspective, the present article broadens this proposal by using a Bayesian approach (based on Markov Chain Monte Carlo (MCMC) methods) to deal with the imprecision resulting from Victora et al.'s model. In order to avoid invalid estimated prevalence values which can occur with the original linear model, truncation or a logit transformation of the prevalences are suggested. The Bayesian approach is illustrated using a community study as an example. Imprecision arising from methodological complexities in the community study design, such as multi-stage sampling and clustering, is easily handled within the Bayesian framework by introducing a hierarchical or multilevel model structure. Since growth deficit was also evaluated in the community study, the article may also serve to validate the procedure proposed by Victora et al.

Key words Anthropometry; Nutritional Surveillance; Statistical Model; Bayes Theorem; Markov chain Monte Carlo Method

Resumo Victora et al. (1998) propuseram o uso de estimativas de prevalência de baixo peso para idade para a estimação de déficit de altura para idade em crianças brasileiras, em virtude da necessidade de simplificar métodos usados em programas de saúde comunitária. Este artigo tenta aprofundar o referido estudo ao propor uma abordagem Bayesiana com base no método de Simulação Estocástica via Cadeia de Markov (SEvCM), para lidar com questões de imprecisão ligadas à modelagem de estimação do déficit de estatura. Para evitar valores inválidos de prevalência estimados pelo modelo linear sugerido originalmente, propõem-se duas alternativas: um truncamento dos valores que extrapolam os limites plausíveis de prevalência ou uma transformação logito das prevalências. A abordagem Bayesiana é ilustrada com um exemplo de um estudo comunitário. Imprecisões oriundas da complexidade do desenho desse estudo também são contornadas com a abordagem Bayesiana, ao se introduzir uma estrutura hierárquica ou multinível. Já que o déficit de crescimento foi efetivamente observado no exemplo, o artigo também serve como instância de validação para o procedimento proposto por Victora et al.

Palavras-chave Antropometria; Vigilância Nutricional; Análise Estatística; Teorema de Bayes; Simulação Estocástica via Cadeia de Markov

Introduction

In a recent article, Victora et al. (1998) proposed the use of low weight-for-age (LWA) prevalence to estimate the prevalence of low height-for-age (LHA). Both indices (weight-for-age and height-for-age) are obtained by comparing the data with reference curves (NCHS, 1978) by sex and age. Values below the -2 z-score cut-off point were used to define deficits.

Focusing on the Brazilian context, Victora et al. (1998) suggest the use of the model

$$[\text{Prev.LHA}] = 0.74 + 2.74 [\text{Prev.LWA}] - 0.03 [\text{Prev. LHA}]^2$$

to estimate the prevalence of LHA, obtained from 38 anthropometric surveys conducted in the country. Justifying this procedure, the authors emphasized the need to simplify methods for use in the context of community health programs, which are precisely those involving the most marginalized segments of the population and which are known to demand more health evaluation and intervention. Underlying the proposal for simplification is the fact that anthropometry requires a certain level of instrumental sophistication when one wishes to achieve minimally acceptable accuracy for decision-making.

However, it is important to note that the regression coefficients in Victora et al.'s model are subject to statistical uncertainty, since they arise from a regression model based on only 38 "data points" (surveys). This additional imprecision should be built in to the final estimate in order to improve the estimation of LHA.

A second issue that must be dealt with concerns the structural specification of the model suggested by Victora et al. (1998). It is possible for their model to predict invalid values for prevalence of LHA – both negative and above 100% – since it is linear with identity link function. Even though in the original article Victora et al. (1998) focused mainly on point estimates, so that this detail did not seem important, this issue becomes overt when the stochastic nature of the estimates are taken into consideration as proposed in the present article.

This article therefore aims to expand on the proposal presented by Victora et al. (1998). Taking the data from a community study as an example (Reichenheim, 1988), we propose a Bayesian approach using Markov chain Monte Carlo (MCMC) methods to obtain prevalence estimates of LHA. This allows one to explicitly incorporate the stochastic nature of the estimates in the predictive model of LHA preva-

lence, in order to make appropriate allowance for their imprecision and thus enable better-informed decisions. This approach also enables spatial and temporal comparisons, which are crucial for community follow-ups of health intervention measures. We also propose either a truncation of values or a re-specification of the model through a logit transformation of the prevalences in order to avoid the problem of invalid estimates of LHA. The latter option not only adequately confines the values to the appropriate distributional space, but also entails a simpler model, eliminating the quadratic term proposed originally by Victora et al. (1988).

An additional feature of the Bayesian approach is that it allows one to simultaneously deal with the difficulties resulting from the use of Victora et al.'s model and with specific methodological issues linked to the complexity of the community study itself. In the present example, the fact that a multi-stage sampling strategy has been employed merits special attention. Since the child is the unit of analysis, two potential sources of imprecision deriving from staging were considered: the census enumeration district (primary sampling unit) and the household (secondary sampling unit). Due to the peculiar characteristics of the community at hand – dispersed and diffuse intra-community health differential, without a well-demarcated geographic pattern – one finds practically all of the variability at the child and household levels. This justifies a modeling approach that incorporate only the clustering effect of the latter. In the current study this is achieved by way of introducing a random effect component to the structural sub-model of the main statistical model.

Taking advantage of the fact that LHA was actually observed in the community study used as an example, another objective of this article is to validate the procedure for using LWA to estimate LHA. We therefore compare the prevalence of LHA estimated by the current procedure with that actually found in the community study, for both the total population and select sub-groups.

The rest of the article is organized as follows. Following a brief introduction to the Bayesian modeling approach in the section on methods, we formally present the various models considered in the analysis and the related modeling procedure. Annotated results are presented in the ensuing section, where we also emphasize the graphical form of the presentation of the findings. In the final section we briefly discuss the possibilities and prospects of the Bayesian method using MCMC in the context of the article.

Methods

Presenting the models considered here and the respective analytical procedures pre-supposes some knowledge of the modeling process from a Bayesian perspective. In short, this involves the following steps (Spiegelhalter et al., 1995a). First, it is necessary to construct a complete probability model, specifying the structural component of the statistical model and the parametric form of the direct relationships between these quantities (sampling model). Having specified the relationships between all the quantities (nodes) in the system, one needs to provide the values for some of these nodes: specifically for the nodes representing the observed data, and for the nodes representing parameters of the Bayesian prior distributions. These priors may be diffuse (vague) or informative, as wished or necessary. It is often convenient to represent such full probability models in a graphical form: an example is given in Figure 1, and discussed in the section “The Direct Acyclical Graph concerning the complete model (F)”.

The model is then implemented using simulation procedures based on Gibbs or Metropolis-Hastings sampling or other MCMC algorithms to generate samples from the posterior distribution of each unknown parameter in the model (Casella & George, 1992; Gilks, 1995). The adequacy of the estimation process must be checked by monitoring the outputs of simulations to ensure mixing and convergence. The latter relates to whether or not the sampler actually reaches the target posterior distribution. Non-convergence can lead to incorrect parameter estimates. Mixing refers to the range of values in the underlying posterior distribution that are actually “visited” (i.e. simulated) by the sampler. If mixing is poor, the sampler may not cover all the sampling space of the target distribution, leading to underestimated standard errors and credibility intervals. Mixing and convergence can be scrutinized through graphic visualization of the simulated chains and calculation of appropriate diagnostics (Cowles & Carlin, 1996).

To complete the process, after evaluating the model's fit through residuals and other indicators (Aitkin, 1991; Kass & Raftery, 1995; Spiegelhalter et al., 1998a), one proceeds to inference based on the posterior distributions of the respective parameters (e.g., means, medians, centiles, standard errors etc.). A more extensive introduction to the topic can be found in Spiegelhalter et al. (1995b) and the application examples accompanying the reference

(Spiegelhalter et al., 1995c, 1995d). For a more in-depth view, we recommend Gelman et al. (1995), the other chapters in Gilks et al. (1995) in addition to that quoted above, and Gamerman (1997).

We now present the model used in the current article. The analytic procedures are covered in the next section.

The model, its components, and alternatives

The model at hand comprises two sub-models. The first sub-model refers to the modeling of LWA prevalence using data from the community survey. The second refers to the modeling of LHA prevalence estimates, i.e., to the model that generates the coefficients used to convert estimates of LWA prevalence into LHA prevalence. We thus start by presenting the LWA sub-model. This will include the specification of priors and the transformation of cluster-specific (CS) regression coefficients (which are directly estimated in the modeling process) into population-average (PA) coefficients (which are necessary for prediction in the context of this analysis). We then describe the various sub-models for estimating LHA prevalence. Six sub-models are contemplated, ranging from the one proposed by Victora et al. (1998) to a completely re-specified sub-model. In the course of the presentation we make several observations justifying the use of the various alternatives.

Sub-model 1 related to the prevalence estimation of low weight-for-age

If one is interested in the prevalence of growth deficit (LHA) in the total population, as well as in some sub-groups relevant from a health point of view (e.g., children with low birthweight or high maternal parity), one must first estimate the overall LWA prevalence and LWA prevalence for the various strata. The outcome variable ([LWA]) and the variables identifying sub-groups are all binary (defining low birthweight [LBW] in the conventional way as < 2500 g, and high parity [PAR] as ≥ 4 children per mother) and so sub-model 1 may be written algebraically as a binary logistic regression model as follows:

$$[LHA]_{ij} \sim \text{Bernoulli}(\pi_{ij}) \quad (1a)$$

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1[LBW]_{ij} + \beta_2[PAR]_i + u_i \quad (1b)$$

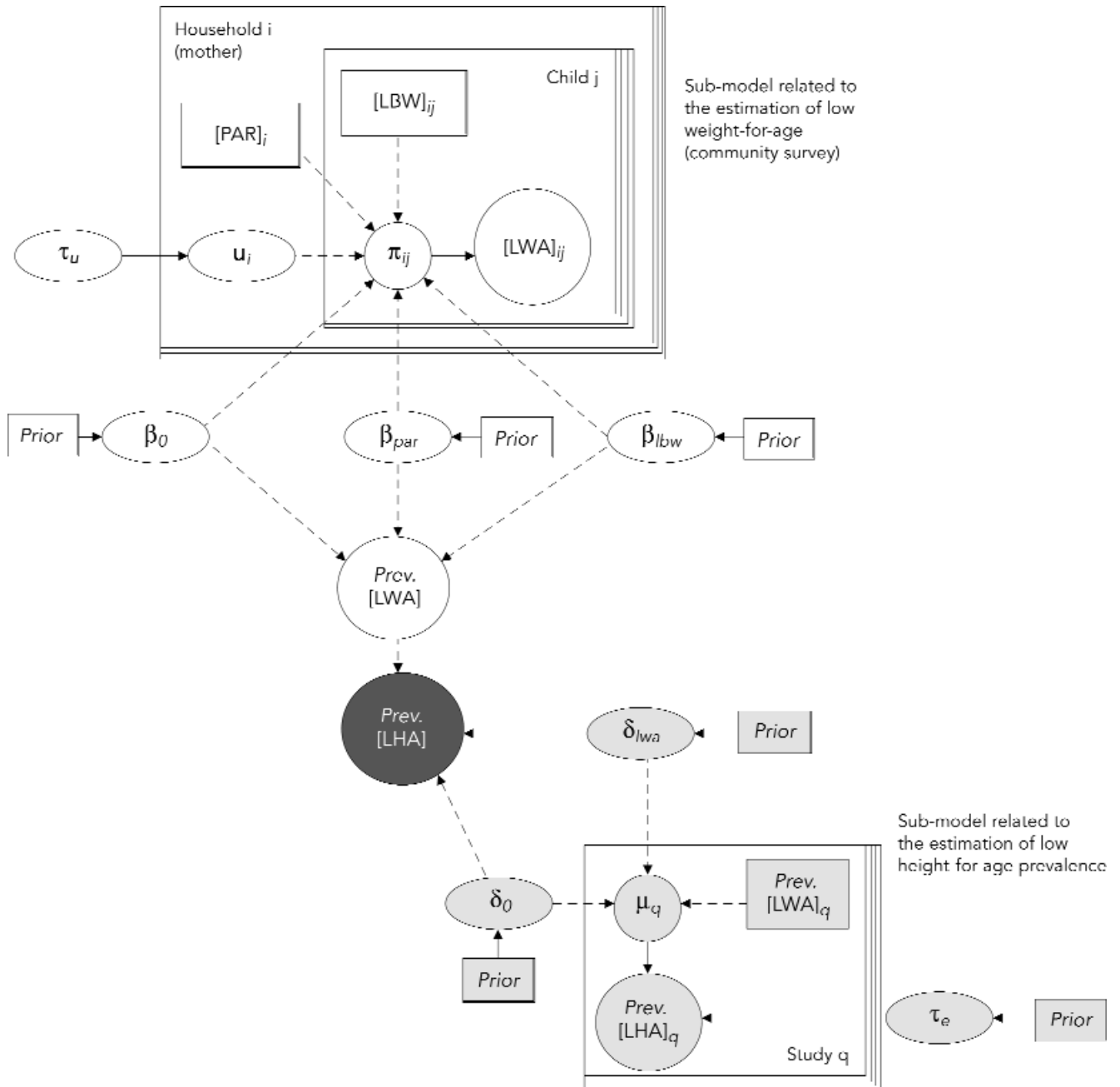
$$u_i \sim \text{Normal}(0, \tau_u) \quad (1c)$$

$$\beta_k \sim \text{Normal}(0, 0.001) \quad (1d)$$

$$\tau_u \sim \text{Gamma}(3, 1) \quad (1e)$$

Figure 1

Direct Acyclic Graph referring to model F (sub-model 1 + sub-model 2F).



The index i refers to the household/mother and j refers to the child. A random effect term u_i is introduced to deal with the design effect (clustering by household/mother). Note that this component is specified as a Gaussian distribution with mean 0 and precision τ_u , the inverse of the variance ($\tau_u = \sigma_u^{-2}$). Equations (1d) and (1e) define, respectively, the priors of the $k = (1, \dots, 3)$ regression coefficients (β_0, β_1 and β_2), and the precision τ_u of the random effect term. The specification of the hyperparameters of the gamma distribution are based on the arguments provided by Smith et al. (1995). Bearing in mind that the model is of the logistic form, this prior states the belief that, amongst children of different households/mothers but with otherwise identical observed covariates, we expect a spread of roughly one order of magnitude in the odds of a positive outcome. A summary of the rationale leading to the actual values for the gamma distribution can be found on page 39 in Spiegelhalter et al. (1995b). On the variance scale (τ^{-1}), this prior places most of its probability mass between 0 and 1, with negligible mass on values above 2; from a substantive point of view, random effect variances much larger than 2 are unlikely in this type of model (Best & Spiegelhalter, 1995). Thus, although it can be characterized as informative, this prior can be considered diffuse over the range of values with epidemiological meaning. It must be pointed out that in the specific case of the community survey used as an example, sampling was self-weighted and thus eliminates the need to incorporate sampling expansion weights.

Sub-model 1 also entails the transformation of the cluster-specific (CS) coefficients that are directly estimated in the analysis into population-average (PA) coefficients (Neuhaus et al., 1991; Diggle et al., 1994; Goldstein & Rasbash, 1996;). The CS coefficients are conditional on the cluster random effect, while what is of most interest here are the marginal coefficients referring to the entire population (or strata thereof). These PA coefficients are necessary for carrying out the predictions presented later in this article. There are various methods for obtaining PA coefficients from CS coefficients. Some require integration over the distribution of the random effects (Zeger et al., 1988). An alternative method proposed by Zeger et al. (1988) simply uses a deterministic approximation to the relationship between the CS and PA coefficients. Since the results are practically indistinguishable, and with a view towards computing efficiency, only the latter is used. The approximation proposed by Zeger et al. (1988) is

$$\beta_k^* \approx \beta_k (1 + 0.346 \sigma_u^2)^{-1/2} \quad (2)$$

where β_k^* represent the desired PA estimations, β_k represent the CS regression coefficients in the random effects model (eq. 1a – 1e), and $\sigma_u^2 = \tau_u^{-1}$ is the random effect variance.

In preparation for the procedure to estimate LHA prevalence, it is necessary to obtain the prevalence estimates of LWA, both for the total and by sub-group. To calculate the latter, one uses the simple conversion

$$[\text{Prev.LWA}]_g = (1 + \exp \{-(\beta_0^* + \beta_1^* [I_{LBW}] + \beta_2^* [I_{PAR}])\})^{-1} \times 100 \quad (3)$$

where the index g indicates sub-groups whose prevalence of LWA are being estimated. $[I_{LBW}]$ and $[I_{PAR}]$ represent the sub-group indicators and can thus take the values 0 or 1, according to the stratum that is being calculated. Note that PA coefficients are used.

To obtain the estimate of the overall prevalence, one calculates

$$[\text{Prev.LWA}]_{ij} = (1 + \exp \{-(\beta_0^* + \beta_1^* [LBW]_{ij} + \beta_2^* [PAR]_{ij})\})^{-1} \times 100 \quad (4)$$

for each child j from household (mother) i , followed by the mean $[\text{Prev.LWA}]_{(\bullet)} = \sum_{ij} [\text{Prev.LWA}]_{ij} / n$. Note that here equation (4) refers to each child ij and thus the individual values of the covariates are used. In calculating the means, n is the total number of children in the community study.

Sub-model 2 related to the prevalence estimation of low height-for-age

Having obtained estimates of the true prevalence of LWA, one proceeds to estimate the prevalence of LHA. A summary of the sub-models can be found in Table 1.

In sub-model 2A, one applies the equation proposed by Victora et al. (1998) to the point estimates of $[\text{Prev.LWA}]_{(\bullet)}$ and $[\text{Prev.LWA}]_g$ obtained from the analysis of the community survey. Generically, sub-model 2A is

$$[\text{Prev.LHA}] = \alpha_0 + \alpha_1 [\text{Prev.LWA}] + \alpha_2 [\text{Prev.LWA}]^2 \quad (5)$$

where, from Victora et al. (1998), $\alpha_0 = 0.743$, $\alpha_1 = 2.342$, and $\alpha_2 = -0.029$.

In sub-model 2B, one directly and simultaneously uses the point estimates of the coefficients provided by Victora et al. (1998). Although equation (5) is formally the same, here the prediction of LHA is carried out at the same time as the analysis of the community data

Table 1

Summary of sub-models 2 for estimating prevalence of low height-for-age.

Characteristics	Estimation sub-models					
	2A	2B	2C	2D	2E	2F
Estimation parallel to analysis of community data		√	√	√	√	√
Estimated value of [Prev.LHA] treated stochastically						
No	√					
Yes: information on model's precision						
Excluded		√				
Included			√	√	√	√
Parameters estimated directly from surveys (anthropom.)				√	√	√
Model adequate to confine zero values estimated by						
Truncation					√	
Specification						√

used to estimate LWA. Since the regression coefficients in sub-model 2B are treated as known (i.e. without allowing for imprecision), the subsequent distribution of [Prev.LHA] is only dependent on the variability stemming from the community survey data in estimating [Prev.LWA].

In 2C, the regression coefficients are treated stochastically and assumed to arise from a Gaussian distribution with mean and standard deviation given, respectively, by the point estimates and standard errors reported by Victora et al from their analysis of the data from the 38 anthropometric surveys. Formally,

$$[\text{Prev.LHA}] = \delta_0 + \delta_1[\text{Prev.LWA}] + \delta_2[\text{Prev.LWA}]^2 \quad (6a)$$

$$\delta_0 \sim \text{Normal}(\alpha_0, \tau_0) \quad (6b)$$

$$\delta_1 \sim \text{Normal}(\alpha_1, \tau_1) \quad (6c)$$

$$\delta_2 \sim \text{Normal}(\alpha_2, \tau_2) \quad (6d)$$

where α_0 , α_1 , and α_2 are the same as for model 2A (equation 5) and $\tau_0 = 0.272$, $\tau_1 = 9.595$, and $\tau_2 = 8548.1$ are the precisions, defined as the inverse square of the respective standard errors of the coefficients reported by Victora et al. (1998). In this model, imprecision arising from the anthropometric surveys is taken into account and transmitted to the prevalence estimate of LHA, which are already dependent on the imprecision of the LWA estimates from the community survey itself. Note that the external information (from Victora et al.'s analysis) is explicitly used as a priori knowledge.

In sub-model 2D, estimation of LHA prevalence is carried out by simultaneously using

the data from both the community survey and the 38 external anthropometric surveys. Here, the coefficients are estimated by

$$\mu_q = \delta_0 + \delta_1[\text{Prev.LWA}]_q + \delta_2[\text{Prev.LWA}]_q^2 \quad (7)$$

$$[\text{Prev.LHA}]_q \sim \text{Normal}(\mu_q, \psi) \quad (8)$$

Index q indicates the prevalence values from each of the 38 anthropometric studies and ψ represents the precision of a normal distribution. Vague priors are assigned to the parameters δ_k ($k = 1, K, 3$) and ψ :

$$\delta_k \sim \text{Normal}(0, 0.001) \quad (9a)$$

$$\psi \sim \text{Gamma}(0.001, 0.001) \quad (9b)$$

The estimates δ_k , are then used to predict [Prev.LHA] for the community study according to equation (6a), where [Prev.LWA] are the prevalence estimates for LWA in the community study, obtained using sub-model 1.

Sub-model 2E is similar to 2D but includes a simple truncation to constrain the predicted LHA prevalence estimates to the range (zero, 100%). In the case of the analysis at hand, this has been accomplished by setting any negative estimation to zero thus avoiding inappropriate values originating from the specification of the sub-model as in 2B through 2D (linear with identity link function).

With the same purpose, model 2F has been re-specified by way of a logit transformation of the prevalences. This avoids arbitrary truncation, since values under 0 and over 100% cannot even be estimated. In a preliminary exploration

of this model, we found that the quadratic term suggested by Victora et al. (1998) is no longer necessary (the quadratic term on the logit scale had a p-value of 0.61 and the model excluding this term still gave a value of $R^2 = 0.88$). In this new sub-model, the 2 coefficients $\delta_0^{(L)}$ and $\delta_1^{(L)}$ are estimated on the basis of the prevalences (LWA and LHA) from the 38 surveys transformed into the respective logits $[P/(1 - P)]$. By also transforming the prevalences of LWA (aggregate and by sub-group) estimated in sub-model 1 to the same scale, one obtains $[\text{Prev.LHA}]^{(L)}$ through

$$[\text{Prev.LHA}]^{(L)} = \delta_0^{(L)} + \delta_1^{(L)}[\text{Prev.LWA}]^{(L)} \quad (10)$$

with priors for $\delta_k^{(L)}$ as in (9a). Thereafter, one transforms back to original scale, obtaining the values of interest:

$$[\text{Prev.LHA}] = (1 + \exp \{-[\text{Prev.LWA}]^{(L)}\})^{-1} \quad (11)$$

As for the objective of validating the LHA prevalence estimation procedure, one also needs the prevalence estimates of LHA from the community study data. For this purpose we used the same structure of sub-model 1, but with [LHA] rather than [LWA]. This modeling thus includes the random component taking the design effect into account.

The Direct Acyclic Graph for the complete model (F)

For purposes of illustration, model F (sub-model 1 + sub-model 2F) is shown in Figure 1. Such representation is based on the formal graphical modeling ideas discussed by Spiegelhalter (1998). Each quantity in the model appears in a node in the graph. For clarity we use rectangular and oval nodes to distinguish data and known constants from unobserved variables, respectively, although both types of node are treated as random variables from a Bayesian perspective. Circular nodes represent deterministic functions. Repetitive structures like children nested within households (mothers) are shown as stacked "sheets". Arrows linking the nodes indicate direct dependencies between the quantities in the model. The solid and dotted lines distinguish between stochastic and deterministic relationships, respectively.

Analyses procedures

Estimation of parameters was carried out in the Bayesian software WinBUGS 1.2, which uses MCMC sampling algorithms. The code referring to the model presented in Figure 1 is shown in Table 2.

To improve estimability or computing efficiency, some authors recommend re-parameterizations of the model (Gilks & Roberts, 1995; Spiegelhalter et al., 1995b). Here, the model has been re-parameterized to center the random effects about the population intercept (rather than about zero), and to center the covariate about its mean. This improves numerical stability and convergence, but does not alter the interpretation of the model.

Mixing was evaluated visually by means of dynamic traces monitored in WinBUGS (Spiegelhalter et al., 1998b). Among many suggested in the literature (Best et al., 1994; Gelman, 1995), convergence was monitored by the potential scale reduction statistic, a diagnostic procedure proposed by Gelman and Rubin (Gelman & Rubin, 1992; Gelman, 1995) and implemented in WinBUGS and the CODA suite of S-functions (Best et al., 1994). For this purpose, we carried out 2 independent simulations of 12,000 updates for the B and C models and 2 simulations of 15,000 updates for the D through F models, beginning with overdispersed initial values as indicated in the literature (Gelman, 1995; Spiegelhalter et al., 1995b). Based on recommended criteria (Gelman & Rubin, 1992; Best et al., 1994), it became clear that convergence and good mixing had been achieved beyond the first 6,000 and 7,500 updates, respectively. The results presented here are based on chains of 12,000 (models B and C) and 15,000 (models D through F) iterations, obtained by pooling the second half of the 2 chains used for the diagnosis.

Results and comments

Estimation models coefficients

Values of the coefficients referring to all six sub-models 2 are found in Table 3. Note that the point estimates and credibility limits (based on 2.5th and 97.5th centiles of the posterior distribution) obtained by MCMC are the same as those estimated traditionally (maximum likelihood), to within Monte Carlo (simulation) error.

To provide a better picture, traces of the chains and distributions for the 2 coefficients in sub-model 2F are found in Figure 2. Note

Table 2

WinBugs code for implementing model F.

```

{
# SUB-MODEL 1

  for (j in 1:J) {
    D[j] ~ dbern(p.D[j]);
    log(p.D[j]) <-      beta0[hhold[j]] +
                      beta.PAR*(PAR[hhold[j]]-0.1974638) +
                      beta.bw2*(LBW[j]-0.0996377);
  }

# Priors

  for (i in 1:I) {
    beta0[i] ~ dnorm(beta0.C, tau.beta0);
  }

beta0.C ~ dnorm(0.0.1.0E-3);           # centered intercept
beta.int <- beta0.C-(beta.PAR*0.1974638)- # return to intercept
           (beta.LBW*0.0996377);       # original
beta.PAR ~ dnorm(0.0. 1.0E-3) ;       # coef. for PAR
beta.LBW ~ dnorm(0.0. 1.0E-3) ;       # coef. for LBW
tau.beta0 ~ dgamma(3,1) ;             # precision
sigma2.beta0 <- pow(tau.beta0. -1) ;   # variance (random effect)
sigma.beta0 <- sqrt(sigma2.beta0) ;    # d.p.

# Transformation of CS coefficient in PA (Zeger et al., 1988)

beta.int.PA <- beta.int/sqrt(((0.346*sigma2.beta0)+1)) ;
beta.PAR.PA <- beta.PAR/sqrt(((0.346*sigma2.beta0)+1)) ;
beta.LBW.PA <- beta.LBW/sqrt(((0.346*sigma2.beta0)+1)) ;

### PREVALENCES AND PREDICTIONS

# sub-model 2F

  for (n in 1:N) {
    prob.LWA.Vict[n] <- LWA.Vict[n]/100 ;
    LWA.Vict.lgt[n] <- log(prob.LWA.Vict[n]) ;
    prob.LHA.Vict[n] <- LHA.Vict[n]/100 ;
    LHA.Vict.lgt[n] <- log(prob.LHA.Vict[n]) ;
    mu[n] <- delta.int + delta.1 * LWA.Vict.lgt[n] ;
    LHA.Vict.lgt[n] ~ dnorm(mu[n], tau.Vict) ;
  }

# Priors for sub-model 2

delta.int ~ dnorm(0. 1.0E-4);
delta.1 ~ dnorm(0. 1.0E-4);
tau.Vict ~ dgamma(1.0E-3, 1.0E-3);

# [Prev.LWA] predicted => prop. D per subgr k

```

Table 2 (continued)

WinBugs code for implementing model F.

```

for (c in 1:C) {
  prob.D.PA[j] <- (1/(1 + (exp(-(beta.int.PA +
    beta.PAR.PA*(PAR[hhold[j]])) +
    beta.LBW.PA*(LBW[j])))))) ;
  prev.D.PA[j] <- prob.D.PA[j]*100 ;

# Conversion using log transf.

prob.D.PA.lgt[j] <- log(prob.D.PA[j]) ;
prob.stunting.PA.lgt[j] <- delta.int + delta.1*prob.D.PA.lgt[j] ;
prev.stunting.PA[j] <- (1 / (1 + exp(-prob.stunting.PA.lgt[j]))) * 100 ;
}

# Total prevalences

prev.D.PA.bar <- mean(prev.D.PA[])
prev.stunting.PA.bar <- mean(prev.stunting.PA[])

# Prevalence in sub-group PAR=1, LBW=0

prob.D.PA.10 <- (1/(1 + (exp(-(beta.int.PA +
  beta.PAR.PA*(1) +
  beta.LBW.PA*(0)))))) ;
prev.D.PA.10 <- prob.D.PA.10*100 ;

# Conversion using log transf.

prob.D.PA.10.lgt <- log(prob.D.PA.10) ;
prob.stunting.PA.10.lgt <- delta.int + delta.1*prob.D.PA.10.lgt ;
prev.stunting.PA.10 <- (1 / (1 + exp(-prob.stunting.PA.10.lgt))) * 100 ;

... Same for sub-groups [PAR=0. LBW=1] , [PAR=0. LBW=0], and [PAR=1, LBW=1]

}

```

that mixing and convergence is adequate, as indicated in the section on methods.

Prevalences of low weight-for-age and low height-for-age

As shown in Table 4, the point prevalence estimates and respective credibility limits for LWA found in the various models are also quite similar. This was to be expected, since sub-model 1 is common to all of them.

The prevalence estimates of LHA are found in Table 5. In general, the point estimates proved to be quite consistent across all models. Slight

differences can easily be attributed to Monte Carlo errors.

However, subtle differences appear when one moves beyond the point estimates. As noted in the section “Sub-models 2 related to the prevalence estimation of low height-for-age”, models that use MCMC (B – F) allow for more appropriate handling of statistical uncertainty. In this respect, the 95% interval estimates for prevalence of LHA under the different models are not equal, either nominally or in terms of adequacy. Intervals in model B are consistently narrower than in the others, capturing quite well the fact that only the imprecision from the

Table 3

Regression coefficients of sub-models 2 – means and centiles 2.5 and 97.5.

Estimation sub-models	δ_0	δ_1	δ_2
Conventional estimation ¹	0.743 (-3.155 / 4.641)	0.341 (1.685 / 2.997)	-0.0294 (-0.0514 / -0.0075)
2C	0.737 (-2.980 / 4.440)	0.340 (1.720 / 2.970)	-0.0296 (-0.0509 / -0.0081)
2D	0.758 (-2.950 / 4.590)	0.340 (1.680 / 2.940)	-0.0293 (-0.0497 / -0.0069)
2E	0.771 (-3.210 / 4.680)	0.330 (1.660 / 3.0)	-0.0291 (-0.0515 / -0.0065)
Conventional estimation ²	0.691 (0.397 / 0.984)	0.930 (0.810 / 1.049)	–
2F	0.685 (0.394 / 0.976)	0.928 (0.809 / 1.050)	–

¹ Values estimated in Stata (StataCorp, 1997) by the maximum likelihood method, based on data from the 38 surveys and using the conversion equation suggested by Victora et al. (1998).

² Same using model with logit-transformed prevalence.

community survey data is encompassed in the former model.

Among the models that incorporate information on the imprecision arising from the LHA estimation procedure, model C is the one with the greatest uncertainty as shown by the wider credibility intervals.

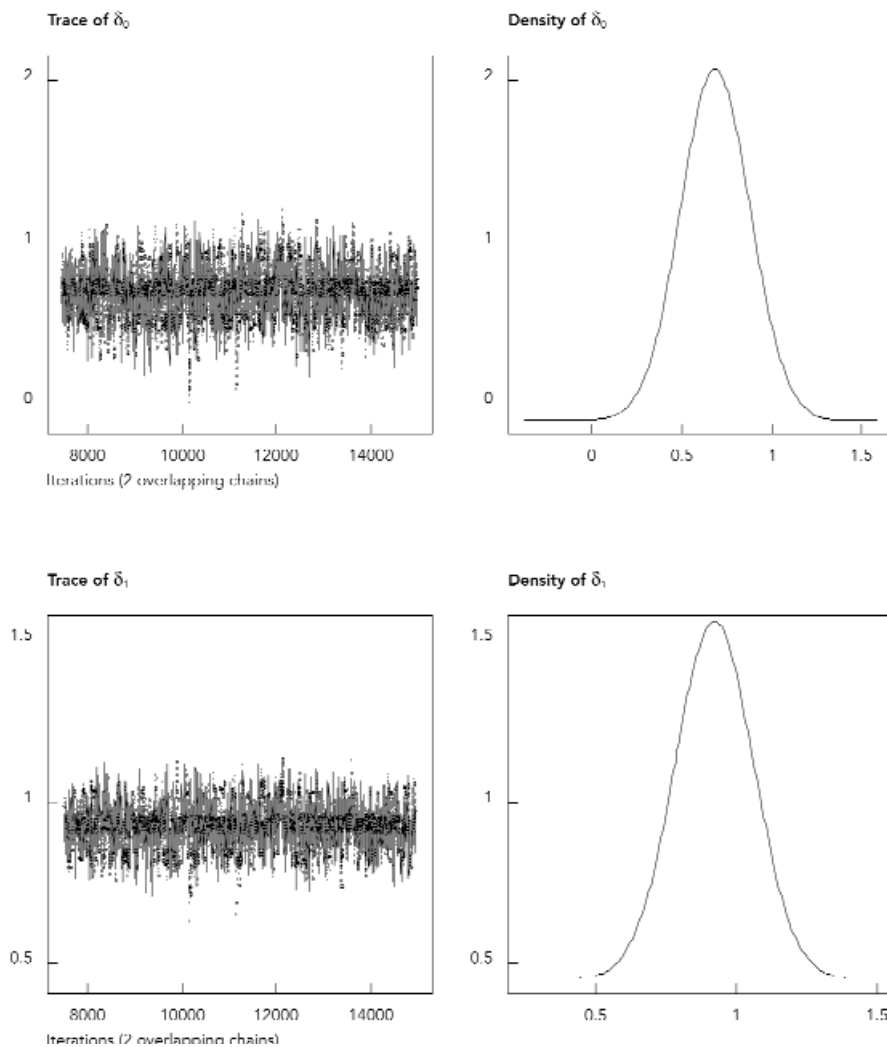
Overall, the stochastic linear models (B – E) appear well-behaved with quite similar point estimates for prevalence ranges above 5%. However, it is in the low range prevalences that one notices problems. Model C, for example, allows for the sampling of negative prevalence values! Although one can argue that from a Bayesian perspective the negative values found specifically for the sub-group [PAR = 0, LBW = 0] would already be in a scarcely plausible range (i.e., in the left tail of the parameter distribution), the same might not occur in other situations with even lower prevalence or perhaps, in studies with a smaller sample size. Indeed, the same could have happened with model D, although apparently it did not. One must point out that the value of 0.2% found in the 2.5th centile is fortuitously positive. Also for model E, nearly 2.5% of the simulated values were negative, as suggested by the lower credibility limit of 0.01 for the sub-group [PAR = 0, LBW = 0]. By arbitrarily truncating the negative values to zero, that of centile 2.5 becomes artificially close to zero (0.01%).

Model F, which uses a logit transformation of the prevalences in the estimation process, is able to deal satisfactorily with this issue. In addition, the 95% credibility intervals for predicted LHA prevalence are narrower, partly as a result of the logit transformation which penalizes extreme values of prevalence more heavily than the linear model.

Drawing attention to the comparisons between the prevalence values for LHA obtained by the various models and those observed in the community survey, similarities are apparent in the range of 5% to 15/20%. Compare, for instance, the values obtained in A or F with the observed values for the total population and sub-groups [PAR = 1, LBW = 0] and [PAR = 0, LBW = 1]. Yet, in ranges outside those limits, a hasty appraisal, based exclusively on point estimates, could lead to a spurious conclusion. Models A and F estimate, respectively, prevalences of 4.1% and 3.9% for sub-group [PAR = 0, LBW = 0] and 40.6% and 39.1% for sub-group [PAR = 1, LBW = 1]. Comparing estimated and observed values, there seem to be some problems at both extremes (respectively, under- and over-estimation by about 30%). However, a closer look, taking the dispersion of the estimators into account, reveals quite a different picture. Clearly, the credibility intervals encompass the point estimates observed in the survey, indicating that there are no major validity problems in

Figure 2

Traces and distributions of sub-model 2F coefficients.



the proposed method. The “true” values in fact fall within the expect limits, especially in light of the models that seek to incorporate the various sources of error.

Posterior distributions of prevalence values estimated by model F

Figures 3a – 3e show the full potential of the Bayesian perspective in the current context, presenting the 5 posterior distributions estimated by model F (sub-model 1 + sub-model 2F). The graphs emphasize the various regions

of plausibility for the LHA prevalence values. Thus, for example, one can infer that the prevalence of LHA for the entire child population in the target community has a substantial possibility of falling between 6% and 8%. As discussed previously, appraisals of this sort are relevant when applied to inter- or intra-community comparisons over time. For example, a drop in the prevalence of growth deficit (LHA) from 7.2% (the point estimate found using model F) to 5.8% in 3 years could give the false impression of effectiveness of an intervention program. However, from a stochastic perspec-

Table 4

Prevalence of malnutrition using the weight-for-age index (≤ -2 z-score) – means and centiles 2.5 and 97.5 – according to models containing and not containing coupled sub-models 2 (B through F).

Estimation models	Total Population	Prevalence (%)			
		Sub-group			
		PAR=1 LBW=0	PAR=0 LBW=1	PAR=0 LBW=0	PAR=1 LBW=1
Estimation of W/A without estimation of H/A ¹	3.23 (1.96 / 4.78)	6.24 (2.69 / 11.30)	6.74 (2.17 / 14.10)	1.48 (0.63 / 2.79)	23.80 (9.03 / 44.80)
B	3.19 (1.99 / 4.78)	6.14 (2.63 / 11.20)	6.69 (2.16 / 13.90)	1.45 (0.65 / 2.81)	23.70 (8.84 / 44.10)
C	3.21 (1.94 / 4.77)	6.15 (2.62 / 11.30)	6.79 (2.20 / 14.10)	1.46 (0.57 / 2.77)	23.90 (9.05 / 44.40)
D	3.26 (1.99 / 4.86)	6.25 (2.65 / 11.40)	6.84 (2.20 / 14.10)	1.50 (0.61 / 2.84)	23.80 (9.07 / 44.00)
E	3.23 (1.96 / 4.81)	6.20 (2.62 / 11.40)	6.81 (2.18 / 14.10)	1.48 (0.58 / 2.81)	23.80 (9.01 / 43.80)
F	3.22 (1.99 / 4.75)	6.18 (2.64 / 11.30)	6.78 (2.23 / 14.00)	1.47 (0.65 / 2.73)	23.70 (8.96 / 44.10)

¹Values estimated by MCMC (WinBUGS) using sub-model 1 without coupled sub-model 2.

Table 5

Prevalence of growth deficit according to the height-for-age index (≤ -2 z-score) – estimated by models A through F and that observed in the community study – means and centiles 2.5 and 97.5.

Estimation models	Total Population	Prevalence (%)			
		Sub-group			
		PAR=1 LBW=0	PAR=0 LBW=1	PAR=0 LBW=0	PAR=1 LBW=1
A	8.01	14.23	15.21	4.15	40.06
B	7.34 (5.04 / 10.40)	13.80 (6.69 / 23.20)	14.80 (5.66 / 27.50)	4.05 (2.24 / 7.07)	36.90 (19.10 / 46.30)
C	7.40 (2.54 / 12.80)	13.90 (5.22 / 25.60)	15.00 (4.58 - 30.50)	4.08 (-0.31 / 8.83)	37.40 (14.40 / 71.80)
D	7.51 (3.76 / 11.40)	14.10 (6.43 / 23.70)	15.10 (5.45 / 28.00)	4.17 (0.20 / 8.26)	37.50 (19.40 / 53.00)
E	7.45 (3.63 / 11.30)	14.00 (6.29 / 23.60)	15.00 (5.38 / 28.00)	4.15 (0.01 / 8.28)	37.40 (19.30 / 53.20)
F	7.20 (4.65 / 10.30)	13.60 (6.48 / 22.80)	14.60 (5.55 / 27.10)	3.86 (1.77 / 6.79)	39.10 (18.70 / 62.00)
Observed Prevalence¹	8.58 (6.48 / 11.00)	14.00 (8.23 / 21.00)	13.60 (6.47 / 23.50)	6.15 (4.14 / 8.63)	27.70 (13.80 / 45.30)

¹Values estimated by MCMC (WinBUGS) by sub-model 1, using LHA variable observed in the community.

tive, contemplating the plausibility regions for the “true” estimate would reveal a broad overlapping of the central regions in the two distributions. Subsequent decisions regarding health interventions would obviously be completely different.

It is important to highlight that the plausibility regions shown in Figure 3 are based on a model in which the two main sources of imprecision in the data have been addressed. Had the imprecision pertaining to the estimation of coefficients from the anthropometric surveys been ignored, this plausibility range would inevitably decrease, leading to a false sense of security. This issue is particularly relevant in relation to sub-group [PAR = 1, LBW = 1], in which the scarcity of information in the community data allows the information from sub-model 2 to dominate over the likelihood in the resulting posterior distribution of LHW. This is evident in Table 4 (last column), where the credibility intervals increase by some 38% from model B to E

Final remarks

One should not fail to restate the underlying motives in the proposal by Victora et. al (Victora et al., 1998). The use of simplified, easily accessible, yet sufficiently accurate methodologies merits interest and further development. The current study fully endorses this view and has hopefully moved forward in attempting to deal with certain issues not addressed or foreseen previously. For example, model F appropriately incorporates the stochastic nature of prevalence estimates, dealing with various sources of imprecision, and offers a solution to the problem of estimating negative prevalence values of LHA, whilst producing a more parsimonious model.

Such issues, in addition to the possibility of providing transparency in the results with the aim of assisting decision-making processes, are positive points in the approach proposed by the current study. Yet, it is worth pointing out that the advantages go beyond the specific problems dealt with here. More complex analytical issues such as, for example, measurement errors (Richardson & Gilks, 1993) and non-ignorable missing data (Best et al., 1996) – possibly found in the various sources of data considered here – can also be easily accommodated and modeled simultaneously.

At first glance, there is a contradiction between the perspective involving simplification and the procedure presented here. A possible

Figure 3

Distribution of growth deficit prevalences estimated by model F.

Figure 3a

Aggregate.

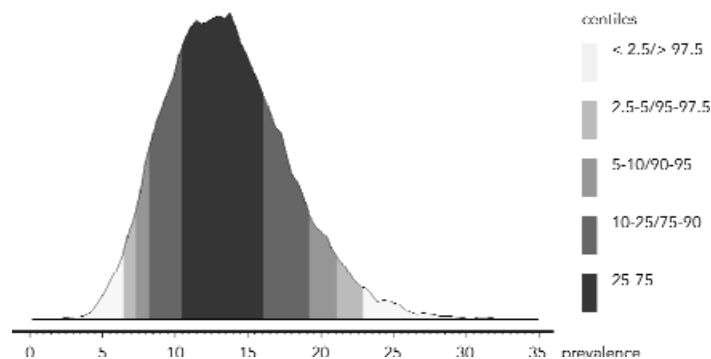
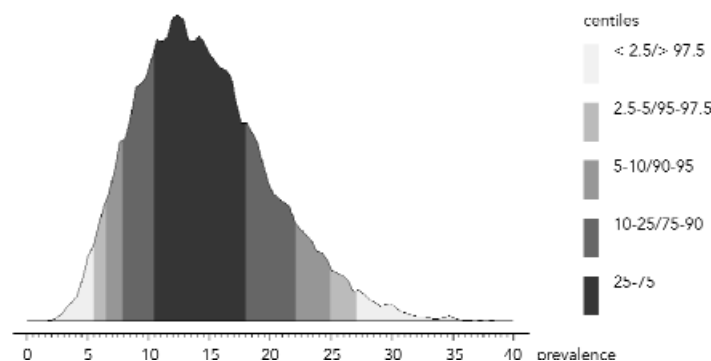


Figure 3b

[PAR = 1, LBW = 0]



objection to adopting a Bayesian approach using MCMC would be the apparent intricacies involved, as opposed to the available expertise in the context where it is meant to be used (e.g., community health programs). Another relates to limited access to hardware and software required to perform the analyses. In fact, the two critiques are acceptable if the proposal is narrowly conceived in terms of complete implementation at the local level, taking the technology's current state of the art into account. We realize that performing the analyses discussed in this article requires some statistical and programming knowledge and, more importantly, computing time and capacity.

Figure 3

Distribution of growth deficit prevalences estimated by model F.

Figure 3c

[PAR = 0, LBW = 1]

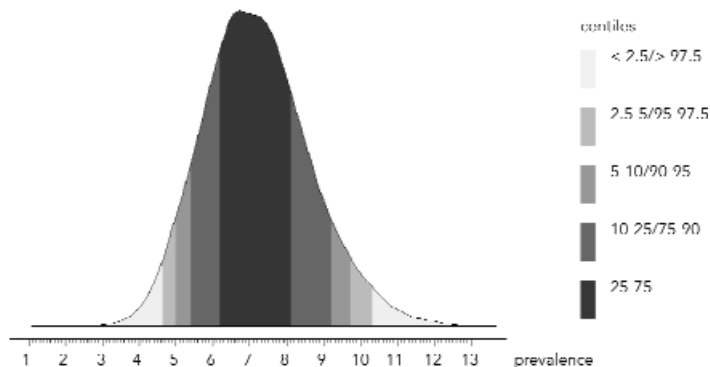


Figure 3d

[PAR = 0, LBW = 0]

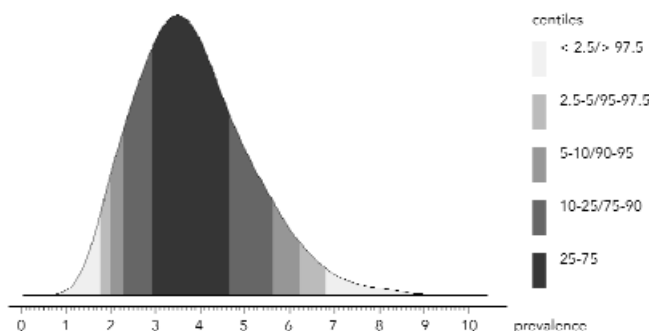
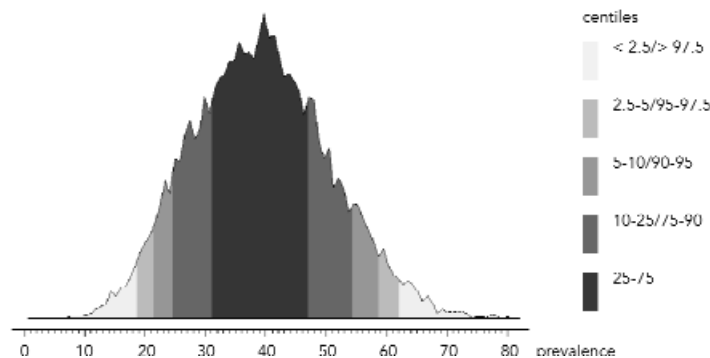


Figure 3e

[PAR = 1, LBW = 1]



However, such objections merit qualification. Computing problems as obstacles tend to decrease gradually with the introduction of more advanced technologies. In a not-so-distant future, computers with parallel processing capabilities should be fully accessible to the public, making computer-intensive methods routine. Suffice it to recall that a logistic regression analysis, now easily performed anywhere by an epidemiologist or public health specialist with minimum training, used to take special programming and a long time to run less than 20 years ago. There is no reason why computer-intensive methods like the Bayesian approach using MCMC should not follow a similar course.

Although there are some software already available (e. g., BUGS 6.0 available on URL <http://www.mrc-bsu.cam.ac.uk/bugs/>; or WinBUGS 1.2 (Spiegelhalter et al., 1998b); or HUGIN (Jensen, 1996)), development will tend to grow in this scenario of expanding computing capacity. And with diversity, more specific and increasingly user-friendly programs may emerge.

The issue of expertise is a more delicate one. No matter how available user-friendly softwares become, the method will always require *ad hoc* programming, essentially because of the flexibility it seeks to provide. However, to overcome this difficulty, one can conceive of intermediate (interface) software that could make the outputs completely transparent and “palatable” to the end user, as witnessed by Figures 3a – 3e. One thus maintains the perspective of using simplified and easily accessible methodologies, even though the underlying procedures are not necessarily so. Such projects will have much to contribute to the establishment of a Bayesian approach in the field of collective health (Etzioni & Kadane, 1995; Landrum & Normand, 1999).

Acknowledgments

The authors are thankful to Cesar Victora for kindly making the data on the 38 surveys available. The authors are also grateful to Antônio Ponce de Leon and Cristiano Fernandes for their valuable comments. MER was partially supported by the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ, grants E-26/171.223/98 and E-26/150.893/99 and CNPq, grant 300234/94-5.

References

- AITKIN, M., 1991. Posterior Bayes factor. *Journal of the Royal Statistical Society, B*, 154:111-142.
- BEST, N. G.; COWLES, M. K. & VINES, K., 1994. *CODA. Convergence Diagnosis and Output Analysis Software for Gibbs. Samples Produced by Bugs Language, Version 0.3*. Cambridge: Medical Research Council Biostatistics Unit.
- BEST, N. G. & SPIEGELHALTER, D. J., 1995. *When is a Random Effect not a Random Effect?* Cambridge: Medical Research Council Biostatistics Unit.
- BEST, N. G.; SPIEGELHALTER, D. J.; THOMAS, A. & BRAYNE, C. E. G., 1996. Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society, A*, 159:323-342.
- CASELLA, G. & GEORGE, E. I., 1992. Explaining the Gibbs sampler. *American Statistician*, 46:167-174.
- COWLES, M. K. & CARLIN, B. P., 1996. Markov chain Monte-Carlo convergence diagnostics – A comparative review. *Journal of the American Statistical Association*, 91:883-904.
- DIGGLE, P. J.; LIANG, K.-Y. & ZEGER, S. L., 1994. *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
- ETZIONI, R. D. & KADANE, J. B., 1995. Bayesian statistical methods in public health and medicine. *Annual Reviews in Public Health*, 16:23-41.
- GAMERMAN, D., 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall.
- GELMAN, A., 1995. Inference and monitoring convergence. In: *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson & D. J. Spiegelhalter, eds.), pp. 131-143, London: Chapman and Hall.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S. & RUBIN, D. B., 1995. *Bayesian Data Analysis*. London: Chapman and Hall.
- GELMAN, A. & RUBIN, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457-472.
- GILKS, W. R., 1995. Full conditional distribution. In: *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson & D. J. Spiegelhalter, eds.), pp. 75-88, London: Chapman and Hall.
- GILKS, W. R.; RICHARDSON, S. & SPIEGELHALTER, D. J., 1995. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- GILKS, W. R. & ROBERTS, G. O., 1995. Strategies for improving MCMC. In: *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson, & D. J. Spiegelhalter, eds.), pp. 89-114, London: Chapman and Hall.
- GOLDSTEIN, H. & RASBASH, J., 1996. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, A*, 159:505-513.
- JENSEN, F. V., 1996. *An introduction to Bayesian Networks*. London: UCL Press.
- KASS, R. E. & RAFTERY, A., 1995. Bayes factor and model uncertainty. *Journal of the American Statistical Society*, 90:773-795.
- LANDRUM, M. B. & NORMAND, S.-L. T., 1999. Applying Bayesian ideas to the development of medical guidelines. *Statistical in Medicine*, 18:117-137.
- NEUHAUS, J. M.; KALBFLEISCH, D. J. & HAUCK, W. W., 1991. A comparison of cluster specific and population-averaged approaches for analysing correlated data. *International Statistical Review*, 59:25-35.
- REICHENHEIM, M. E., 1988. *Child Health in an Urban Context: Risk Factors in a Squatter Settlement of Rio de Janeiro*. PhD Thesis, London: University of London.
- RICHARDSON, S. & GILKS, W. R., 1993. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology*, 138:430-442.
- SMITH, T. C.; SPIEGELHALTER, D. J. & THOMAS, A., 1995. Bayesian graphical modelling applied to random effects in meta-analysis. *Statistics in Medicine*, 14:2685-2699.
- SPIEGELHALTER, D. J., 1998. Bayesian graphical modelling: A case-study in monitoring health outcomes. *Journal of the Royal Statistical Society, C*, 47:115-133.
- SPIEGELHALTER, D. J.; BEST, N. G. & CARLIN, B. P., 1998a. *Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrary Complex Models*. Cambridge: Medical Research Council Biostatistics Unit/University of Cambridge.
- SPIEGELHALTER, D. J.; BEST, N. G.; GILKS, W. R. & INSKIP, H., 1995a. Hepatitis B: A case study in MCMC methods. In: *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson & D. J. Spiegelhalter, eds.), pp. 21-43, London: Chapman and Hall.
- SPIEGELHALTER, D. J.; THOMAS, A. & BEST, N. G., 1995b. *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5*. Cambridge: Medical Research Council Biostatistics Unit.
- SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G. & GILKS, W., 1995c. *BUGS 0.5 Examples, Volume 1*. Cambridge: Medical Research Council Biostatistics Unit.
- SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G. & GILKS, W., 1995d. *BUGS 0.5 Examples, Volume 2*. Cambridge: Medical Research Council Biostatistics Unit.
- SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G. & GILKS, W., 1998b. *WinBUGS (software): Bayesian Inference Using Gibbs Sampling, Version 1.2*. Cambridge: Medical Research Council Biostatistics Unit/Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine at St. Mary's Hospital.
- VICTORA, C. G.; GIGANTE, D. P.; BARROS, A. J. D.; MONTEIRO, C. A. & ONIS, M., 1998. Estimativa de prevalência de déficit de altura/idade a partir da prevalência de déficit de peso/idade em crianças brasileiras. *Revista de Saúde Pública*, 32: 321-327.
- ZEGER, S. L.; LIANG, K.-Y. & ALBERT, P. S., 1988. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44:1049-1060.