POPULATION, SPACE AND PLACE
Popul. Space Place **15**, 461–473 (2009)
Published online 16 December 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/psp.524

Geo-additive Models of Childhood Undernutrition in Three Sub-Saharan African Countries

Ngianga-Bakwin Kandala,[1,*] Ludwig Fahrmeir,[2] Stephan Klasen[3] and Jan Priebe[3]
[1]*Clinical Sciences Research Institute, Warwick Medical School, University of Warwick, Coventry CV2 2DX, UK*
[2]*Department of Statistics, University of Munich, Germany*
[3]*Department of Economics, University of Göttingen, Germany*

ABSTRACT

We investigate the geographical and socioeconomic determinants of childhood undernutrition in Malawi, Tanzania and Zambia, three neighbouring countries in southern Africa, using the 1992 Demographic and Health Surveys. In particular, we estimate models of undernutrition jointly for the three countries to explore regional patterns of undernutrition that transcend boundaries, while allowing for country-specific interactions. We use geo-additive regression models to flexibly model the effects of selected socioeconomic covariates and spatial effects. Inference is fully Bayesian based on recent Markov chain Monte Carlo techniques.

While the socioeconomic determinants generally confirm findings from the literature, we find distinct residual spatial patterns that are not explained by the socioeconomic determinants. In particular, there appears to be a belt transcending boundaries and running from southern Tanzania to northeastern Zambia which exhibits much worse undernutrition. These findings have important implications for planning, as well as in the search for left-out variables that might account for these residual spatial patterns. Copyright © 2008 John Wiley & Sons, Ltd.

*Correspondence to: Ngianga-Bakwin Kandala, Clinical Sciences Research Institute, Warwick Medical School, University of Warwick, Coventry CV2 2DX, UK. E-mail: N-B.Kandala@warwick.ac.uk

Received 10 March 2006; revised 20 November 2007; accepted 23 November 2007

Keywords: sub-Saharan Africa; geo-additive models; undernutrition; spatial statistics; semi-parametric Bayesian analysis

INTRODUCTION

Malawi, Tanzania and Zambia are neighbouring low-income countries in southern Africa, all belonging to the poorest countries in the world, with very poor education, health, and human development indicators. They have been affected by years of economic stagnation and decline, with negative per capita income growth rates throughout the 1980s and early 1990s, and have also experienced deterioration in health and education indicators (World Bank, 2000).[1] Chronic undernutrition is a serious problem in all three countries, affecting some 48.7% of children in Malawi, 46.7% in Tanzania, and 39.6% in Zambia. Given the centrality of undernutrition for child well-being, it is critical to understand its determinants.

When modelling the determinants of undernutrition, one can distinguish between immediate, intermediate, and underlying determinants (see UNICEF, 1998). While undernutrition is always immediately related to either insufficient nutrient intake or the inability of the body to absorb nutrients (primarily due to illness), these are themselves caused by problems related to food security, care practices, and the health environment at the household level, which themselves are influenced by the socioeconomic and demographic situation of households and communities (UNICEF, 1998; Smith and Haddad, 1999; Klasen, 2000). In order to capture this complex chain of causation, researchers have either focused on a particular level of causality (e.g. Smith and Haddad, 1999; Pelletier, 1998), have estimated structural equations that address interactions between the different levels (e.g. Guilkey and Riphahn, 1998), have used graphical chain models to assess the causal pathways (Caputo *et al.*, 2003), or have used multilevel modelling techniques (e.g. Nyovani *et al.*, 1999; Harttgen and Misselhorn, 2006).

In this paper, we estimated a model that mainly focuses on factors that are underlying determinants of undernutrition. There is also some discussion in the literature on non-linear effects of some of these covariates, although they are rarely investigated thoroughly (e.g. World Health Organization, 1995; Stephenson, 1999; Smith and Haddad, 2001).

The particular innovations of the paper are three-fold. Firstly, through the use of our empirical methods we are able to investigate non-linear effects more flexibly than most previous work. Secondly, the methods also allow us to investigate the spatial pattern of undernutrition, prior to and after controlling for the socioeconomic covariates. This enables us to determine to what extent the substantial spatial pattern of undernutrition is driven by the socioeconomic factors we were considering. Thirdly, by using data from three adjacent countries in the same year, we are able to investigate the relative importance of country-specific socio-economic factors and government policies vis-à-vis geographical factors that transcend boundaries.

DATA, MEASUREMENT OF UNDERNUTRITION, AND COVARIATES

The data we use are from the 1992 round of the Demographic and Health Surveys (DHS) undertaken in the three sub-Saharan African countries of Malawi, Tanzania and Zambia. The DHSs collect information on a nationally representative sample of women of child-bearing age (15–49). The questionnaire collects socioeconomic indicators for the respondent and her partner, as well as the household she resides in, and then gathers a large amount of information on fertility patterns, health and care practices, health knowledge, and assesses the anthropometric status of all children of these women who were born within the past five years. Data collection and analysis is highly standardised and supported from Macro, Inc. on behalf of the US Agency for International Development. The DHSs have long been considered to be the most reliable sources of demographic, nutrition, and child health and mortality information in developing countries, particularly in Africa where administrative data on these issues are largely incomplete or missing. Unfortunately, the surveys do not generate an income variable, and we therefore rely on a household asset index as a proxy for the income situation of the households which has been found to be quite reliable by Filmer and Pritchett (2001). The 1992 DHS data-sets of Malawi, Tanzania and Zambia are pooled to form one data-set with the same socio-economic, demographic and health characteristics of the household. This is possible because the DHS surveys were carried out in a standardised form, with the same list of socio-economic and demographic characteristics.

The DHS samples for the three countries were drawn through stratified clustered sampling comprising some 18,000 children in 835 clusters. Although not available in the data-set, we were able to obtain the district location of each cluster and can therefore base our spatial analysis on the 156 districts in the three countries.[2]

Undernutrition among children is usually measured by determining the anthropometric status of the child, with most research focusing on children below six years of age (e.g. World Health Organization, 2006). Researchers typically distinguish between three types of undernutrition: wasting, or insufficient weight for height, indicating acute undernutrition; stunting or insufficient height for age, indicating chronic undernutrition; and underweight or insufficient weight for age, which could be a result of both stunting and wasting (e.g. UNICEF, 1998). Wasting, stunting, and underweight for a child i are typically determined using a Z-score which is defined as:

$$Z_i = \frac{AI_i - MAI}{\sigma}$$

where AI refers to the individual anthropometric indicator (e.g. height at a certain age), MAI refers to the median of a reference population, and refers to the standard deviation of the reference population. The reference standard typically

DOI: 10.1002/psp

used for the calculation is the National Center for Health Statistics (NCHS) Center for Disease Control (CDC)/Growth Standard that has been recommended for international use by the World Health Organization (World Health Organization, 1983, 1995).[3]

The children whose Z-scores were below –2 standard deviations (SD) from the median of the reference category are considered as undernourished (stunted, wasted, and underweight, depending on the indicator chosen), while those with Z-scores below –3 are considered severely undernourished. In this paper we focus on stunting, as our covariates were better able to explain chronic than acute undernutrition. We used the Z-score (in a standardised form) as a continuous variable to use the maximum amount of information available in the data-set. The geographical distribution of the standardised Z-scores[4] for the response variable stunting, averaged by district (left) and region (right), is displayed in Fig. 1. It shows distinct spatial patterns of undernutrition. While in southwestern Zambia and northern Tanzania it appears that stunting is lower, there seem to be more areas of high stunting in northeastern Zambia, northern Malawi and southern Tanzania. In addition to local small-area variability, there might also be an underlying smooth spatial component which crosses borders, something we investigate below. The comparison between the left and right panel of Fig. 1 also suggests that it is well worth examining the spatial pattern of undernutrition at a more disaggregated district level, as the regional analysis glosses over important intra-regional differentials.

Figure 2 shows a histogram and kernel density estimates of the distribution of the Z-scores, together with a normal density, with mean and variance estimated from the sample. This gives clear evidence that a Gaussian regression model is a reasonable choice for inference.

Regarding the covariates, we were guided by the previous literature on the subject and the conceptual framework outlined by UNICEF (1998). Among the underlying determinants of chronic undernutrition, we consider socioeconomic factors measured by an asset index, household size, the nutritional status of the mother (measured by her body-mass index, BMI), and access to electricity, health knowledge and care practices measured by mother's education, mother's marital status, birth interval, place of delivery, vaccinations, and access to a radio. We also control for the sex of the child, urban–rural location, and the age of the child. Based on our own prior work as well as other literature (e.g. World Health Organization, 1995; Moradi and Klasen, 2000; Kandala *et al.*, 2001; Kandala, 2002), we investigate a potentially non-linear pattern of

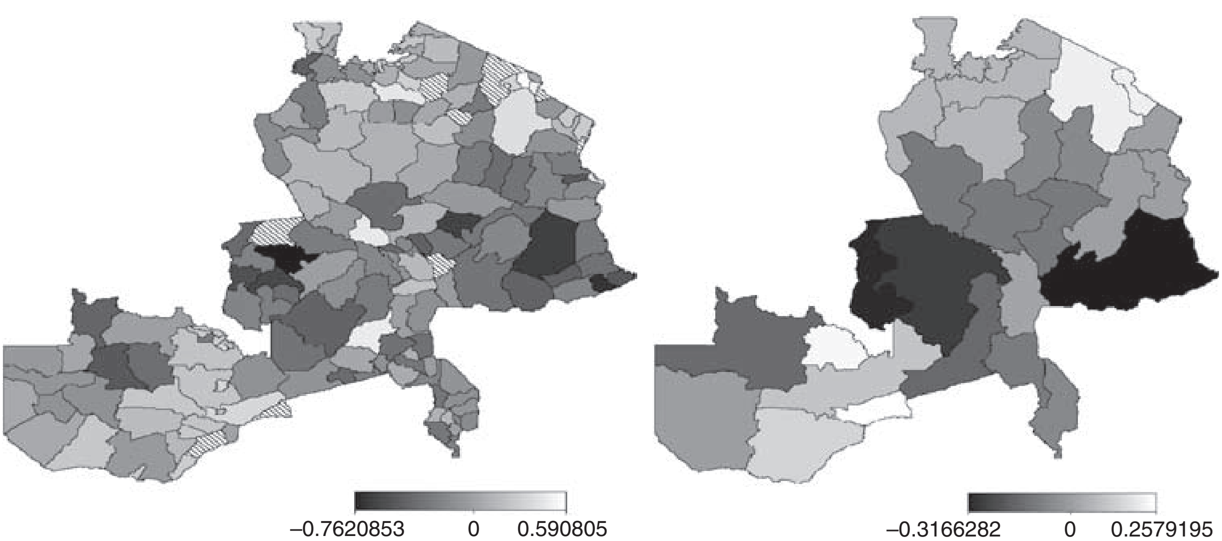


Figure 1. Observed mean Z-score of stunting by districts (left) and regions (right).
Note: Darker areas indicate low Z-scores and thus poor undernutrition, while lighter ones suggest high Z-scores. Moreover, hached areas refer to districts for which no information was available in the data-set.

DOI: 10.1002/psp

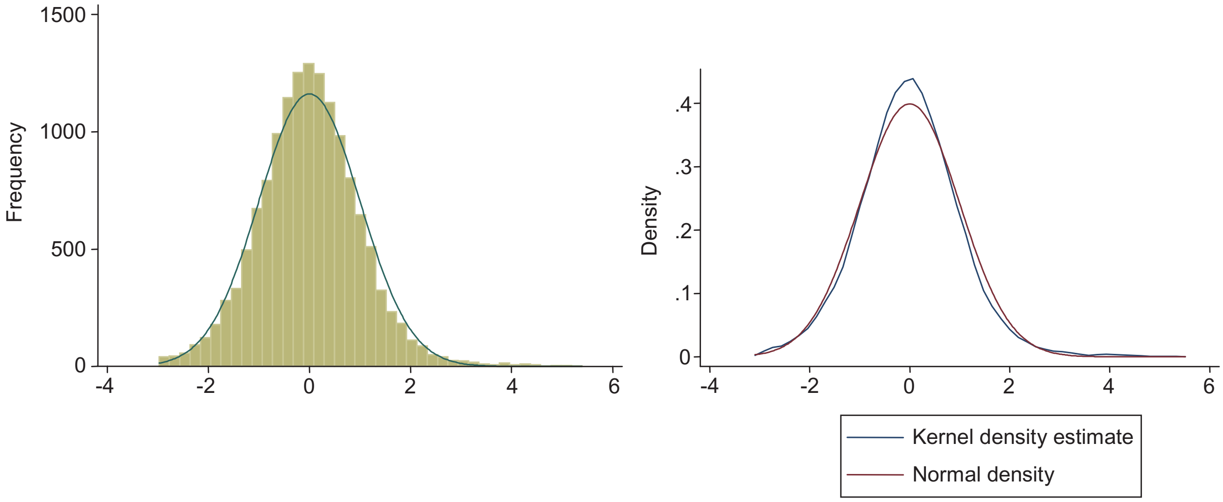


Figure 2. Histogram (left) and kernel density estimates (right) of standardised Z-score for 'stunting'.

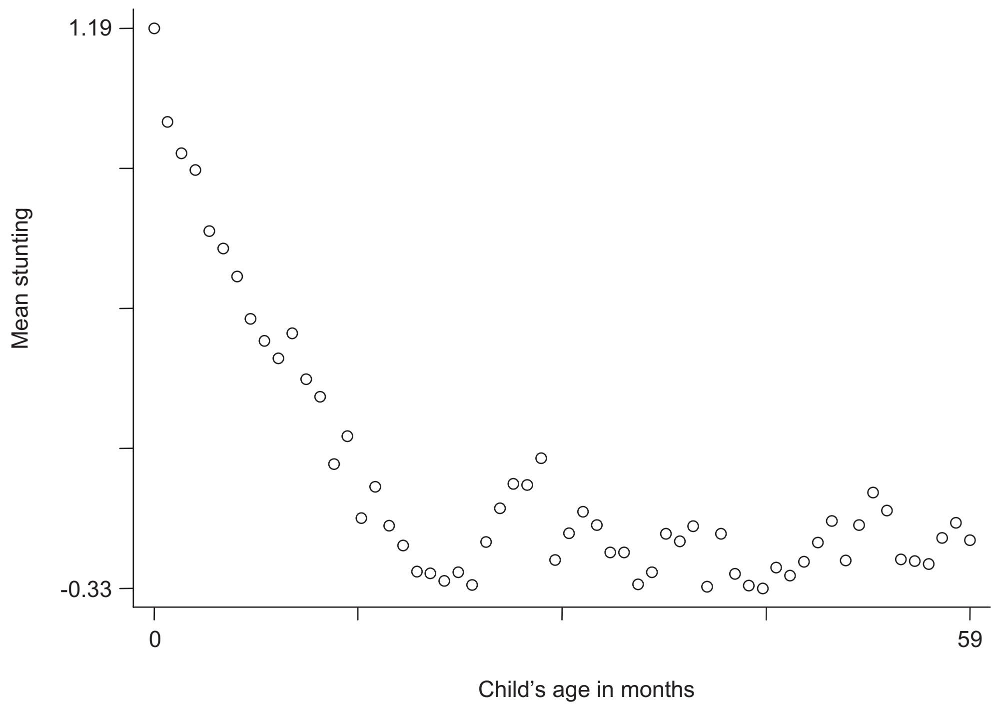


Figure 3. Mean standardised Z-score for stunting by child's age.

effects of the mother's BMI as well as the age pattern on undernutrition. For illustration, the empirical distribution of the stunting Z-score by child's age is shown in Fig. 3. It is obvious that the effect of child's age on the mean Z-score of stunting is non-linear. It will be difficult to model the possibly non-linear effect of such covariates through a parametric functional form, which justifies our use of a flexible semi-parametric model.

Empirical distributions of categorical covariates, together with coding used in the analysis, are given in Table 1. Most of the factors are placed in categories to be compared with previous studies. For instance, household size is split into 'Small households of < than 6 members' (reference), 'Medium households of 6–10 members' and 'Large households of > 10 members' in the three countries. Ownership of consumer items, such as

DOI: 10.1002/psp

Table 1. Overview of covariates.

Factor	Malawi (%)	Tanzania (%)	Zambia (%)	Coding
Residence				
Rural	74.5%	84.3%	57.3%	0: rural, reference
Urban	25.5%	15.7%	42.7%	1: urban
Has radio				
No	54.8%	64.2%	57.2%	0: reference
Yes	45.0%	34.1%	42.4%	1
Has electricity				
No	94.8%	92.8%	80.5%	0: reference
Yes	5.0%	5.6%	19.2%	1
Educational attainment				
No education	41.6%	37.2%	17.9%	0: reference
Incomplete primary	42.8%	18.9%	30.3%	
Complete primary	10.1%	40.6%	32.8%	1 (incl. inc. prim.)
Incomplete secondary	3.5%	3.0%	15.3%	
Completed secondary	1.7%	0.1%	2.1%	2 (incl. inc. & higher)
Higher	0.2%	0.2%	1.5%	
Sex of child				
Female	49.3%	49.9%	49.9%	0: reference
Male	50.7%	50.1%	50.1%	1: male
Child's place of delivery				
Born at home	62.5%	50.2%	49.6%	0: reference
Born in hospital	37.5%	49.8%	60.4%	1: hospital
Current marital status				
Single mothers	11.5%	15.3%	15.2%	1: single
Married women	88.5%	84.7%	84.8%	0: reference
Child received vaccination				
No	22.1%	24.1%	27.18%	0: reference
Yes	77.5%	75.9%	72.8%	1
Asset index				
High socioeconomic status	31.5%	23.4%	50.4%	0: reference
Low socioeconomic status	24.3%	38.9%	32.4%	1: low
Medium socioeconomic status	44.2%	37.8%	17.2%	2: medium
Birth interval				
Short: ≤24 months	23.85%	21.34%	22.84%	0: reference
Long: >24 months	76.15%	78.66%	77.16%	1: long birth
Size of household				
Small size: <6 members	46.72%	32.40	33.45	0: reference
Medium size: 6–10 members	47.23%	50.37	53.94	1
Large size: >10 members	6.05%	17.23	12.61	2
Mean of BMI	21.96	21.75	21.96	Metrical
District	32	62	62	Spatial covariate

a radio or car, as well as characteristics of the dwelling such as floor or roof type, toilet facilities and water source, are items that measure poverty in these settings, and the World Bank and others have used these items to generate an asset index using Principal Components Analysis (PCA) (Filmer and Pritchett, 2001). We use the first principal component derived from the pooled data from the three countries to obtain the index for each household. We sort children by the asset index and establish cut-off values for percentiles of the population. We then refer to the bottom

DOI: 10.1002/psp

third as 'low socioeconomic status', the next third as 'medium socioeconomic status', and the top third as 'high socioeconomic status' (see Table 1).[5]

While all three countries do relatively poorly on the reported socioeconomic indicators, there are significant differences between the countries as well. In particular, households in Zambia appear to be better off in terms of access to electricity, radio, and female educational attainment. Income as proxied by the asset index and educational levels are also higher in Zambia (Table 1). This country is also more heavily urbanised than the other two. Malawi and Tanzania are more similar, with Malawi doing somewhat worse on access to electricity. Malawi also has worse educational attainment at the lower levels but slightly higher among the highest levels than Tanzania. Since the effect of education and the asset index might vary across countries, it seems appropriate to test for interactions for these variables.

BAYESIAN GEO-ADDITIVE REGRESSION MODELS

Spatial analyses of undernutrition are often confined to using region-specific dummy variables to capture the spatial dimension. Here, we go a step further by exploring regional patterns of childhood undernutrition and, possibly non-linear, effects of other factors within a simultaneous, coherent regression framework using a geo-additive semi-parametric mixed model. Because the predictor contains usual linear terms, non-linear effects of metric covariates and geographical effects in additive form, such models are also called geo-additive models. Kammann and Wand (2003) proposed this type of model within an empirical Bayesian approach. Here, we apply a fully Bayesian approach as suggested in Fahrmeir and Lang (2001) and Lang and Brezger (2004), which is based on Markov priors and uses Markov Chain Monte Carlo (MCMC) techniques for inference and model checking.

Classical linear regression models of the form:

$$y_i = w_i'\gamma + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \tag{1}$$

for observations (y_i, w_i), $i = 1, \ldots, n$, on a response variable y and a vector w of covariates assume that the mean $E(y_i|w_i)$ can be modelled through a *linear predictor* $w_i'\gamma$. In our application to childhood undernutrition and in many other regression situations, we are facing the following problems. Firstly, for the *continuous covariates* in the data-set, the assumption of a strictly linear effect on the response y may not be appropriate. In our study, such covariates are the child's age (*age*), the mother's age at birth (*mab*), and the mother's body-mass index (*BMI*). Generally, it will be difficult to model the possibly non-linear effect of such covariates through a parametric functional form, which has to be *linear* in the parameters, prior to any data analysis.

Secondly, in addition to usual covariates, geographical small-area information was given in the form of a location variable s, indicating the region, district or community where individuals or units in the sample live or come from. In our study, this geographical information is given by the districts of the three countries concerned. Attempts to include such small-area information using district-specific dummy variables would in our case entail more than 200 dummy variables, and using this approach we would not assess spatial interdependence. The latter problem also cannot be resolved through conventional multilevel modelling using uncorrelated random effects (Goldstein, 1999). It is reasonable to assume that areas close to each other are more similar than areas far apart, so that spatially correlated random effects are required.

To overcome these difficulties, we replace the strictly linear predictor through a *geo-additive predictor*, leading to the *geo-additive regression model*:

$$y_i = f_1(x_{i1}) + \ldots + f_p(x_{ip}) + f_{spat}(s_i) + w_i'\gamma + \varepsilon_i. \tag{2}$$

where $f_1, \ldots, f_p$ are non-linear smooth effects of the metric covariates, and f_{spat} is the effect of the spatial covariate $s_i \in \{1, \ldots, S\}$ labelling the districts in the three countries. Regression models with predictors as in equation (2) are sometimes referred to as geo-additive models. The observation model (2) may be extended by including interaction $f(x)w$ between a continuous covariate x and a binary component of w, say, leading to so-called varying coefficient models, or by adding a non-linear interaction $f_{1,2}(x_1, x_2)$ of two continuous covariates.

In a Bayesian approach, unknown functions f_j and parameters γ as well as the variance parameter σ^2 are considered as random variables and have to be supplemented with appropriate prior assumptions. In the absence of any prior knowledge we assume independent diffuse priors

DOI: 10.1002/psp

$\gamma_j \propto const$, $j = 1, \ldots, r$ for the parameters of fixed effects. Another common choice is highly dispersed Gaussian priors.

Several alternatives are available as smoothness priors for the unknown functions $f_j(x_j)$ (see Fahrmeir and Lang, 2001; Fahrmeir *et al.*, 2004). We use Bayesian P(enalized)-splines, introduced by Eilers and Marx in a frequentist setting. It is assumed that an unknown smooth function $f_j(x_j)$ can be approximated by a polynomial spline of low degree. The usual choices are cubic splines, which are twice continuously differentiable piecewise cubic polynomials defined for a grid of k equally spaced knot p on the relevant interval $[a, b]$ of the x-axis. Such a spline can be written in terms of a linear combination B-spline basis functions $B_m(x)$:

$$f(x) = \sum_{m=1}^{l} \beta_m B_m(x) \tag{3}$$

These basis functions have finite support on four neighbouring intervals of the grid, and are zero elsewhere. A comparably small number of knots (usually between 10 and 40) is chosen to ensure enough flexibility in combination with a roughness penalty based on second-order difference of adjacent B-spline coefficients to guarantee sufficient smoothness of the fitted curves. In our Bayesian approach this corresponds to second order random walks:

$$\beta_m = 2\beta_{m-1} - \beta_{m-2} + u_m, \tag{4}$$

with Gaussian errors $u_m \sim N(0, \tau^2)$. The variance parameter τ^2 controls the amount of smoothness, and is also estimated from the data. More details on Bayesian P-splines can be found in Lang and Brezger (2004). Note that random walks are the special case of B-splines of degree zero.

We now turn our attention to the spatial effects f_{str} and f_{unstr}. For the spatially correlated effect $f_{str}(s)$, $s = 1, \ldots, S$, we choose Markov random field priors, common in spatial statistics (Besag *et al.*, 1991). These priors reflect spatial neighbourhood relationships. For geographical data one usually assumes that two sites or regions s and r are neighbours if they share a common boundary. Then a spatial extension of random walk models leads to the conditional, spatially autoregressive specification:

$$f_{str}(s) \mid f_{str}(r), \quad r \neq s \sim N\left(\sum_{r \in \partial_s} f_{str}(r)/N_s, \tau^2/N_s\right) \tag{5}$$

where N_s is the number of adjacent regions, and $r \in \partial_s$ denotes that region r is a neighbour of region s. Thus the (conditional) mean of $f_{str}(s)$ is an average of function evaluations $f_{str}(s)$ of neighbouring regions. Again the variance τ^2_{str} controls the degree of smoothness.

For a spatially uncorrelated (unstructured) effect f_{unstr} a common assumption is that the parameters $f_{unstr}(s)$ are independent and identically distributed (i.i.d.) Gaussian:

$$f_{unstr}(s) \mid \tau^2_{unstr} \sim N(0, \tau^2_{unstr}) \tag{6}$$

Variance or smoothness parameters τ^2_j, $j = 1, \ldots, p, str, unstr$, are also considered as unknown and estimated simultaneously with corresponding unknown functions f_j. Therefore, hyper-priors are assigned to them in a second stage of the hierarchy by highly dispersed inverse gamma distributions $p(\tau^2_j) \sim IG(a_j, b_j)$ with known hyper-parameters a_j and b_j. For model choice, we routinely used the Deviance Information Criterion (DIC) developed in Spiegelhalter *et al.* (2002) as a measure of fit and model complexity.

RESULTS

Based on previous analyses carried out separately for each country (Kandala *et al.*, 2001), we chose a geo-additive model with interactions between country-effects and educational attainment as well as the asset index. Taking Tanzania as the reference country, we arrived at the model with interaction terms in Zambia and Malawi for the three levels of education and asset indices (see Table 1):

$$y_i = f_1(agc_i) + f_2(bmi_i) + f_{str}(s_i) + f_{unstr}(s_i) + w_i'\gamma + \varepsilon_i \tag{7}$$

The model assumes that $f_1(\ )$, $f_2(\ )$ and f_{str} are non-linear and spatial effects which are the same for all three countries. This was confirmed by prior separate analyses of the non-linear effects in each of the countries, which were found to be remarkably similar. Table 2 contains the fixed effects, and the non-linear effects of BMI and child's age are shown in Fig. 4. In the left-hand map of Fig. 5 we show the mean Z-scores by district, predicted by considering the socioeconomic covariates only; in the right-hand map we then subtract these predicted Z-scores of the left-hand figure from the raw Z-scores to generate the residuals by district that are not

DOI: 10.1002/psp

Table 2. Posterior mean of fixed effects.

Variable	Mean	10% quant.	90% quant.
Constant	–0.14	–0.22	–0.07
Urban	0.15	0.11	0.21
Male	–0.10	–0.12	–0.08
Incomplete primary & complete primary education in Tanzania	0.03	–0.01	0.06
Incomplete & complete secondary and higher in Tanzania	0.35	0.24	0.46
Additive effect of incomplete & complete primary education in Malawi	0.08	–0.04	0.12
Additive effect of incomplete & complete secondary and higher in Malawi	0.07	–0.02	0.13
Additive effect of incomplete & complete primary education in Zambia	–0.003	–0.06	0.05
Additive effect of incomplete & complete secondary and higher in Zambia	–0.17	–0.30	–0.05
Middle household in Tanzania	0.01	–0.02	0.05
Rich household in Tanzania	0.18	0.14	0.23
Additive effect of middle household in Malawi	–0.05	–0.11	–0.01
Additive effect of rich household in Malawi	–0.06	–0.13	0.02
Additive effect of middle household in Zambia	–0.01	–0.07	0.06
Additive effect of rich household in Zambia	–0.02	–0.09	0.05
Single mothers	–0.07	–0.10	–0.04
Long birth interval (>24 months)	0.08	0.06	0.11
Medium household (6–10 members)	0.03	–0.004	0.05
Large household (>10 members)	0.07	0.03	0.09

Note: Left-out categories are rural, female, no education, poor households in Tanzania, married women, short birth-interval, and small households.

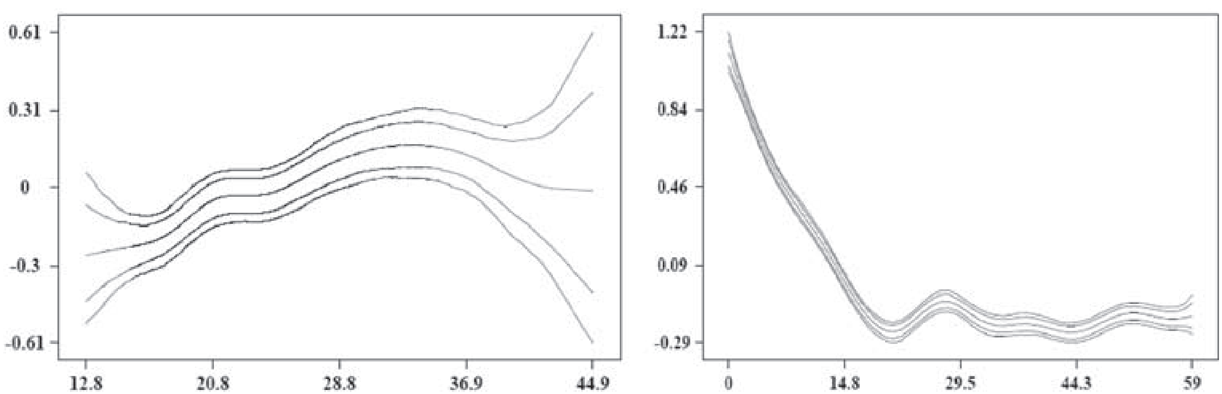
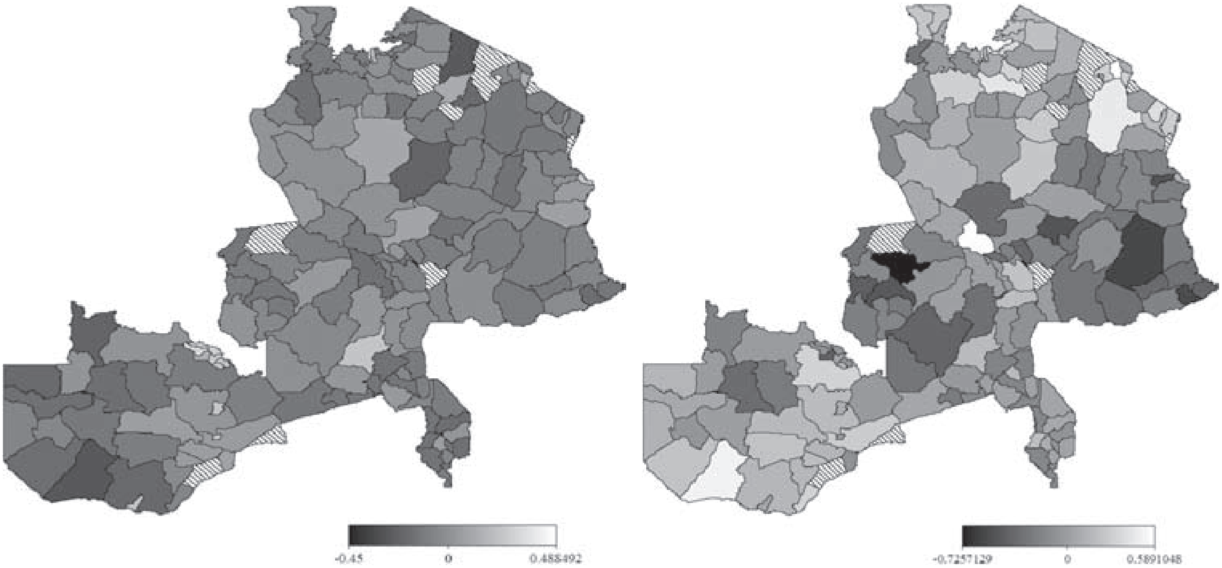


Figure 4. Non-linear effects of mother's body-mass index (left) and child's age (right) on stunting.

explained by the socioeconomic variables. These are then allocated to structured and unstructured effects. The posterior mean estimate of the structured spatial effects f_{str} is shown in the maps of Fig. 6. Figure 7 shows the unstructured spatial effects.[6] In addition, posterior probability maps indicate significance of the spatial effects (white/black = significantly positive/negative effect on the Z-score, grey = not significant). Note that the spatial effects are centred around zero, such that the average over all districts is zero, while the overall level is estimated through the intercept term. Before commenting on the substantive results, it is important to point out this model had the best fit after evaluation of the fit criteria using Deviance Information Criteria (DIC).

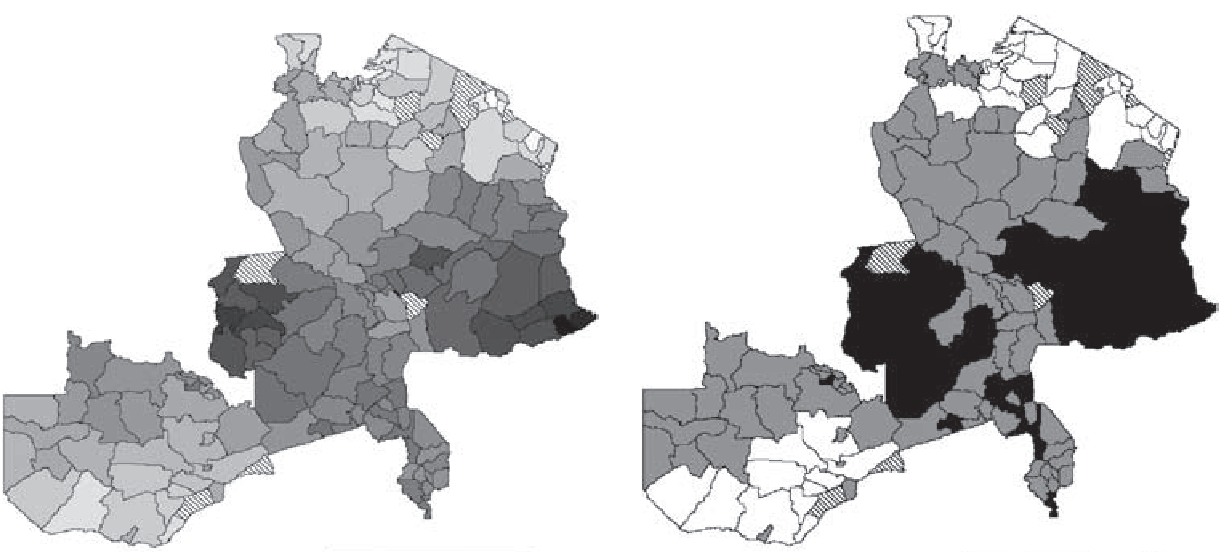
The results for the fixed effects in Table 2 suggest that female children are slightly less stunted, which had also been found in other studies (Klasen, 1996; Kandala *et al.*, 2001, 2006). This is indicated by the fact that the corresponding posterior mean of 0.09 for males is negative and the 10% and 90% quantiles are both negative, indicating that the effect is statistically

DOI: 10.1002/psp



Light – low prevalence of stunting
Dark – high prevalence of stunting

Figure 5. Mean of stunting predicted by the covariates for Model 5 (left) and residual spatial effects of stunting (right).



Light – low prevalence of stunting
Dark – high prevalence of stunting

Black – significant negative effect
Grey – not significant
White – significant positive effect

Figure 6. Structured residual spatial effect of stunting (left) and posterior probabilities (right) of stunting for Model 5.

DOI: 10.1002/psp

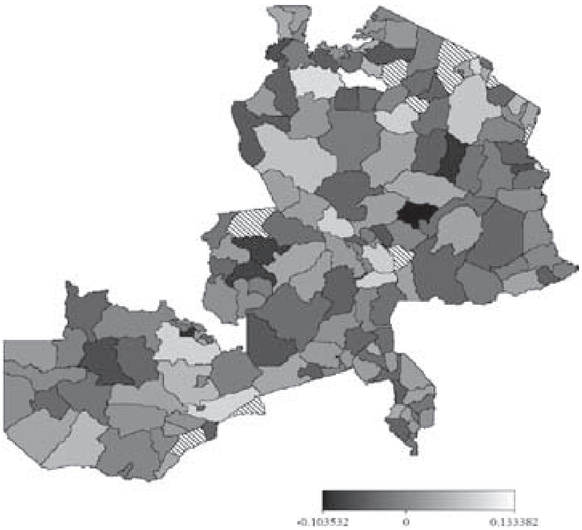


Figure 7. Unstructured spatial effects. See Fig. 6 for key.

significant. It is also the case that children living with both parents (married women) are less stunted than other children, apparently benefiting from the extra care of both parents. Alternatively, couples may benefit from economies of scale for childcare as well as in expenditures (Kandala, 2002; Kandala *et al.*, 2006). Children from low socioeconomic households were, as expected, more stunted than children from high income backgrounds. The prior birth interval matters for the nutritional status of the child. The analysis shows that children born after a long birth interval are better off than other children.

Better nutritional status is observed in households with a large number of adult members. The impact of the household's size should not, however, be over-interpreted, since to some extent it directly mirrors infant mortality. For instance, a household with high mortality risk will remain small. In contrast, a household's size might also reflect its wealth, as a rich household will attract occupants (Kandala, 2002). Again, in a large household, a child might benefit from the help of several adults. Large households may benefit from scale economies in childcare as well as in expenditures. Alternatively, they may have become better at raising children through accumulated experience (Christiansen and Alderman, 2001).

The only country interactions that turned out to be significant and thus were retained in the model are the interactions with mother's education and asset index. Here we find that the positive effect of high levels of mother's education is much smaller in Zambia than in Tanzania. This might be related to the fact that the economic crises of the early 1990s in Zambia also affected more educated groups, who were unable to shield their children from these adverse conditions. Similarly, the positive effect of middle and high income to the household is also much smaller in Malawi and Zambia. It thus appears that these two socioeconomic indicators have a much larger effect in Tanzania than elsewhere.

The left panel of Fig. 4 shows the flexible modelling of the effect of BMI of the mother. Shown are the posterior means together with 80% and 95% pointwise credible intervals. We found the influence to be in the form of an inverse U shape. While the inverse U looks nearly symmetrical, the descending portion exhibited a much larger range in the credible region. This appears quite reasonable, as obesity of the mother (possibly due to a poor-quality diet) is likely to pose less of a risk for the nutritional status of the child as very low BMIs, which suggest acute undernutrition of the mother. The Z-score is highest (and thus stunting lowest) at a BMI of around 30–35.

The right panel of Fig. 4 shows the effect of the child's age on its nutritional status. As suggested by the nutritional literature, we are able to discern the continuous worsening of nutritional status up until about 20 months of age. This deterioration sets in right after birth and continues, more or less linearly, until 20 months. Such an immediate deterioration in nutritional status is not quite as expected, as the literature typically suggests that the worsening is associated with weaning at around 4–6 months. One reason for this unexpected finding could be that, according to the surveys, most parents give their children liquids other than breast milk shortly after birth which might contribute to infections.

After 20 months the effect of age on stunting stabilises at a low level. Through reduced growth and the waning impact of infections, children are apparently able to reach a low-level equilibrium that allows their nutritional status to stabilise. We also see a blip at around 24 months of age. This is picking up the effect of a change in the data-set that makes up the reference standard. Until 24 months, the currently used international reference standard is based on white children in

DOI: 10.1002/psp

the US of high socioeconomic status, while after 24 months it is based on a representative sample of all US children (World Health Organization, 1995). Since the latter sample exhibits worse nutritional status, comparing the Tanzanian children to that sample leads to a sudden improvement in their nutritional status at 24 months. This drawback of the reference standard is one reason for the recent introduction of a new reference standard (World Health Organization, 2006).

Figure 5 (left) shows that the socioeconomic effects are able to explain a fair amount of the spatial variation in undernutrition in the three countries. We calculate that the average residual spatial effect in the right-hand panel of Fig. 5 is about 30% lower than the original spatial effects plotted in Fig. 1, showing that socioeconomic effects explain some but not all of the spatial variation. However, the spatial residuals in the right-hand side of Fig. 5 show that much of the variation in stunting remains to be explained. These spatial effects are then allocated by the model into structured effects which are shown in Fig. 6, and unstructured residual effects in Fig. 7.

Several important findings emerge. Firstly, many of these structured spatial effects are significant as indicated by the right panel of Fig. 6, which shows the posterior probability maps of undernutrition at a 95% credible interval. The districts in black indicate a significant negative spatial effect (more undernutrition), while the districts in white imply a significant positive spatial effect. The rest of the districts (in grey) have no significant effect on undernutrition. Thus we clearly have a pattern of worse nutrition in eastern and northeastern Zambia, central Malawi and southern Tanzania. Conversely, Z-scores are significantly better in northern Tanzania and southwestern Zambia.

Secondly, while these structured effects suggest worse undernutrition in a belt ranging from northern Zambia to southern Tanzania, it is interesting to note that the districts in northern Malawi are not significant components in that belt. Thus while some spatial residuals do spill significantly across borders (e.g. between northern Zambia and central Malawi), some borders do seem to matter in the sense that spatial residuals remain noticeably distinct in the analysis on the two sides of borders. Thirdly, the unstructured spatial effects shown in Fig. 7, while being much smaller and not significant, also display an interesting pattern. While in Tanzania large cities have significantly higher Z-scores, this was not the case in Zambia where some of the large cities in the Copperbelt (the small districts in the central-north of the country) actually have lower Z-scores. This may be related to the effect of the decline in copper production and the impact of general economic decline and structural adjustment policies that have affected urban areas more than rural areas (World Bank, 2000).

The clear structured pattern begs an explanation. None of the socioeconomic variables we tried in addition to the ones mentioned are able to reduce these pronounced spatial effects. One common factor to most of the areas that are negatively affected is that these areas are at comparatively low elevations. This distinction is most noticeable and clear in the south–north divide in Tanzania, but it is also noticeable elsewhere. The difference could well be due to differences in disease prevalence such as malaria, schistosomasis, and other diseases that thrive at lower elevations and were particularly problematic along the Rift Valley. In an exploratory analysis, we compare the spatial pattern of prevalence of fever, diarrhoea, cough, or any of the three illnesses combined with the structured spatial pattern, and found that the spatial distribution of fever (presumably related to malaria) has a fairly close resemblance to the structured spatial effects while the others did not appear to play a significant role. Future work should explore this linkage further. The measure of disease prevalence used here, recall of whether anyone in the household had been ill with fever, cough or diarrhoea in the past two weeks, is less than perfect as it is quite subjective, based on short-term recall, and has considerable noise. Future work needs to address the question of disease environment more closely.

Moreover, the poor nutritional status in northeastern Zambia could additionally be related to poor access to health facilities and the general remoteness of these areas, which are poorly served with transportation links (World Bank, 2000). These issues deserve closer attention and this procedure is merely able to highlight the important spatial patterns of undernutrition without being able to explain them fully.

Quite clearly, the methods used here are able to identify more subtle socioeconomic and spatial influences on undernutrition than reliance on linear models with regional dummy variables. As such, they are useful for diagnostic purposes to identify the need to find additional variables that can account for this spatial structure. Moreover, even if the causes of the spatial structure are not fully explained, one can use this spatial information for poverty mapping and associated planning purposes, which is gaining increasing importance in policy circles that attempt to focus the allocation of public resources to the most deprived sections of the population.

CONCLUSIONS

In this paper we pooled the 1992 Demographic and Health surveys of Malawi, Tanzania and Zambia to model the socioeconomic and spatial determinants of undernutrition. We find strong support for our approach of flexibly modelling some covariates that clearly have non-linear influences and for including a spatial analysis. The spatial analysis shows distinct patterns that point to the influence of omitted variables with strong spatial structure or possibly epidemiological processes that account for this spatial structure.

The maps generated could be used for targeting development efforts or for exploring relationships between welfare indicators and other variables. For example, a mortality or undernutrition map could be overlaid with maps of other types of data, say on poverty, agro-climatic or other environmental characteristics. The visual nature of the maps may highlight unexpected relationships that would be overlooked in a standard regression analysis. These maps are novel tools to help policy-makers achieve Millennium Development Goals (MDGs) for child health in these countries.

ACKNOWLEDGEMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation), and the Collaborative Research Center 386: Analysis of Discrete Structures. We thank participants at seminars in Munich, Tübingen, and at IFPRI in Washington, DC, as well as two anonymous referees, for helpful comments and discussions. We also wish to thank Macro Int. for providing us with the district locations of the clusters in the three DHS data-sets.

NOTES

(1) More recently, they have also been severely affected by the AIDS pandemic.
(2) We cannot use later DHS surveys undertaken in the three countries as they either did not take place in the same year or the district information was not available.
(3) Recently, a new reference standard has been generated from which Z-scores can be calculated. For the purposes of this paper, the use of the new reference standard would not change the qualitative results. For a discussion of the new reference standard, see World Health Organization (2006) and Klasen (2007).
(4) The standardised Z-score, calculated as the actual Z-score minus its median divided by its standard deviation, was used for computational purposes.
(5) Other categorical covariates, such as employment situation of the mother and type of toilet facility, turned out to be non-significant in the preliminary data analysis and are thus omitted.
(6) The unstructured effects are smaller in magnitude and none are significant, so we do not include a significance map with this figure but briefly comment upon them.

REFERENCES

Besag J, York Y, Mollie A. 1991. Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**: 1–59.

Caputo A, Roraita R, Klasen S, Pigeot I. 2003. Undernutrition in Benin: an analysis based on graphical models. *Social Sciences and Medicine* **56**: 1677–1691.

Christiansen L, Alderman H. 2004. *Child Malnutrition in Ethiopia: Can Maternal Knowledge Augment the role of Income?* World Bank.

Eilers PHC, Marx BD. 1996. Flexible smoothing with B-splines and penalties. *Statistical Sciences*. **11**: 89–121.

Fahrmeir L, Kneib T, Lang S. 2004. Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*. **14**: 731–761.

Fahrmeir L, Lang S. 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics* (*JRSS C*) **50**: 201–220.

Filmer D, Pritchett L. 2001. Estimating wealth effects without income of expenditure data-or tears: an application to educational enrollment in states of India. *Demography* **38**: 115–132.

DOI: 10.1002/psp

Goldstein H. 1999. *Multilevel statistical models* (1st Internet Edn). Available at http://multilevel.ioe.ac.uk/index.html.

Guilkey D, Riphahn R. 1998. The determinants of child mortality in the Philippines: estimation of a structural model. *Journal of Development Economics* **56**: 281–305.

Harttgen K, Misselhorn M. 2006. *A multilevel approach to explain child mortality and undernutrition in South Asia and Sub-Saharan Africa*. Mimeograph, University of Göttingen. Available at http://www2.vwl.wiso.uni-goettingen.de/ibero/papers/DB152.pdf

Kammann EE, Wand MP. 2003. Geoadditive models. *Journal of the Royal Statistical Society C* **52**: 1–18.

Kandala N-B. 2002. *Spatial modelling of socio-economic and demographic determinants of childhood undernutrition and mortality in Africa*. PhD thesis, University of Munich, Shaker Verlag.

Kandala N-B, Lang S, Klasen S, Fahrmeir L. 2001. Semiparametric analysis of the socio-demographic determinants of undernutrition in two African countries. *Research in Official Statistics*, *EUROSTAT* **4**: 81–100.

Kandala N-B, Magadi MA, Madise NJ. 2006. An investigation of district spatial variations of childhood diarrhoea and fever in Malawi. *Social Science and Medicine* **62**: 1138–1152.

Klasen S. 1996. Nutrition, health, and mortality in sub-Saharan Africa: is there a gender bias? *Journal of Development Studies* **32**: 913–932.

Klasen S. 2000. *Malnourished and low mortality in South Asia*, *better nourished and dying young in Africa: what can explain this puzzle?* SFB 386 Discussion Paper No. 214. University of Munich. Available at http://www.stat.uni-muenchen.de/sfb386/welcome.html.

Klasen S. 2007. Poverty, undernutrition, and child mortality: some inter-regional puzzles and their implications for research and policy. *Journal of Economic Inequality* (published on-line in March 2007). DOI: 10.1007/s/10888–007–9056-x.

Lang S, Brezger A. 2004. Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**: 183–212.

Moradi A, Klasen S. 2000. *The nutritional status of elites in India*, *Kenya*, *and Zambia: an appropriate guide for developing International Reference Standards for undernutrition?* SFB 386 Discussion Paper No. 217. University of Munich. Available at http://www.stat.uni-muenchen.de/sfb386/welcome.html.

Nyovani JM, Matthews Z, Margetts B. 1999. Heterogeneity of child nutritional status between households: a comparison of six sub-Saharan African countries. *Population Studies* **53**: 331–343.

Smith L, Haddad L. 2001. *The Importance of Women's Status for Child Nutrition in Developing Countries*. IFPRI: Washington, DC.

Spiegelhalter D, Best N, Carlin B, Van der Line A. 2002. Bayesian measures of models complexity and fit. *Journal of the Royal Statistical Society B* **64**: 1–34.

Stephenson C. 1999. Burden of infection on growth failure. *Journal of Nutrition* (Suppl.) 534S–538S.

UNICEF. 1998. *The State of the World's Children*. UNICEF: New York.

World Bank (2000). Profile of poverty in Malawi, 1998. Poverty analysis of the Malawian Integrated Household Survey, 1997–98. Washington, DC: The World Bank.

World Health Organization. 1983. *Measuring Change in Nutritional Status*. WHO: Geneva.

World Health Organization. 1995. *Physical Status: The Use and Interpretation of Anthropometry*. WHO Technical Report Series No. 854. WHO: Geneva.

World Health Organization. 2006. Multicentre Growth Reference Study. *Acta Paediatrica* **450**(Suppl.): 1–87.