Analysis of Childhood Morbidity with Geoadditive Probit and Latent Variable Model: A Case Study for Egypt

Khaled Khatab* and Ludwig Fahrmeir
Institute of Occupational and Social Medicine, Medical Faculty, RWTH Aachen University, Aachen, Germany; Department of Statistics, Ludwig-Maximilians-Munich University, Munich, Germany

Abstract. This work applies geoadditive latent variable models to analyze the impact of risk factors and the spatial effects on the latent, unobservable variable “health status” or “frailty” of a child less than 5 years of age using the 2003 Demographic and Health survey (DHS) data from Egypt. Childhood diseases are a major cause of death of children in the developing world. In developing countries a quarter of infant and childhood mortality is related to childhood disease, particularly to diarrhea. Our case study is based on the 2003 Demographic and Health Survey for Egypt (EDHS). It provides data on the prevalence and treatment of common childhood disease such as diarrhea, cough, and fever, which are seen as symptoms or indicators of children’s health status, causing increased morbidity and mortality. These causes are often associated with a number of risk factors, including inadequate antenatal care, lack of or inadequate vaccination, and environmental factors that affected the health of the child in early years. Various bio-demographic and socioeconomic variables, in this work, we examine the impact of such factors on childhood disease with flexible geoadditive models. These models allow us to analyze usual linear effects of covariates, nonlinear effects of continuous covariates, and small-area regional effects within a unified, semi-parametric Bayesian framework for modeling and inference. As a first step, we use separate geoadditive probit models the binary target variables for diarrhea, cough, and fever using covariate information from EDHS. Based on these results, we then apply recently developed geoadditive latent variable models where the three observable disease variables are taken as indicators for the latent individual variable “health status” or “frailty” of a child. This modeling approach allows us to study the common influence of risk factors on individual frailties of children, thereby automatically accounting for association between diseases as indicators for health status.

INTRODUCTION

Childhood diseases are among the most serious health issues facing developing countries. About half of all childhood deaths are caused by diseases such as pneumonia, diarrheal diseases, malaria, and measles. High infant and childhood mortality rates obviously induce low life expectancy in many developing countries and have severe negative impact on future development. Thus, the investigation of proximate determinants that influence the risk of diseases and the outcome of disease processes is highly important; see some contributed chapters in Mosley and Chen for early references and, for example, recent work by Kandala and Kandala and others. The main objective of this work is to examine the impact of the socioeconomic and bio-demographic factors on childhood disease, including geographic effects as a surrogate for unobserved covariates with spatial information. In our case study, we focus on the analysis for childhood disease in Egypt using data from the 2003 Demographic and Health survey (EDHS), which was jointly sponsored by the United Nations Population Fund Activities (UNPFA) and the U.S. Agency International Development (USAID). One of the main objectives of EDHS is to provide up-to-date information on childhood disease. This tends to assist policy makers and administrators in evaluating and designing programs and improving planning for future interventions in these areas, which in turn should reduce childhood morbidity and childhood mortality as well. We will model the impact of various socioeconomic, public health, and geographic variables on disease of young children in developing countries with Egypt as a case study. Selection of the variables is inspired by the conceptual framework for proximate determinants of childhood morbidity and mortality developed in Mosley and Chen. In practice of course, selection is limited by the set of covariates available in the EDHS data set at hand. Statistical analysis will be based on modern Bayesian approaches, which allow flexibly formulating realistically complex geoadditive regression and latent variable models. In a first step, we analyze the impact of various risk factors on the three diseases diarrhea, cough, and fever through separate geoadditive probit models developed in Fahrmeir and Lang and in Brezger and Lang. In a second step, we use geoadditive latent variable models recently suggested in Raach and in Fahrmeir and Raach. In geoadditive probit latent variable models, the three observable binary disease variables are taken as indicators for the latent individual variable “health status” or “frailty” of a child. This modeling approach allows studying the common influence of risk factors on individual frailties of children, thereby automatically accounting for association between diseases as indicators for health status. Compared with previous results, our approach can provide additional and new insight into childhood morbidity and mortality in developing countries in general and, more specifically, in Egypt.

Previous studies on child disease have focused on various socioeconomic, demographic, or health factors that are available in specific data sets. However, most of these studies have neglected aspects for the spatial effects; see for instance Miller and Hirschhorn and Miller and others. A notable exception is the study for Nigeria in Kandala and others, which uses separate geoadditive probit models for cough, fever, and diarrhea. Studies for Egypt are restricted to a few selected or specific towns and governorates. Also, most articles focused only on diarrhea, using simple statistical analyses (see Abu-Elyazed and others, El-Gilany and Hammad, Langsten and Hill). Our case study differs from these previous works. First, the analysis studies spatial differentials of child disease at a highly disaggregated governorates level using a Bayesian approach for geoadditive models. This allows incorporating covariates...
effects in a flexible semi-parametric way, which is not possible through the usual parametric approaches applied in previous works. Second, a latent variable model (LVM) for health status based on binary disease indicators permits modeling of covariate effects on the latent variables through a flexible geodative predictor. This model gives us the opportunity to study the association or interrelationship between the three types of diseases as indicators for health status. The factor loadings describe the association between the diseases and their impact on the health status of a child.

All computations have been carried out with BayesX Version 1.40, STATA and R Programs using the MCMC package; see Raach and Fahrmeir and Raach. The rest of the work is organized as follows. The Materials and Methods section describes the data set and methods followed by a description of geodative models and latent variable models. The results of the separate models are shown also in the Materials and Methods section. The Discussion section discussed the results of the separate models and the results of the latent variable models are shown in this section. The final remarks included in the Conclusion section.

MATERIALS AND METHODS

This work is based on data available from the 2003 EDHS. The 2003 EDHS uses standard survey instruments to collect data on household members, such as working status and education of mother, sex of child, exposure to media, etc. It also collects household living conditions, such as housing characteristics and information on fertility, mortality and child health from mothers in reproductive ages (15–49). Individual data records were constructed for 6,661 children in Egypt. Each record consists of disease information and the list of covariates that could affect the child’s health. The EDHS data permit attributing child morbidity to three specific diseases, namely diarrhea, fever, and cough, observed in the last 2 weeks before the survey. In developing countries, including Egypt, these three specific diseases are often symptoms for diseases such as malaria, acute respiratory infections, stomach infections, etc., that in turn are responsible for increased child mortality. Table 1 shows an overview of the three common diseases in Egypt in the last 2 weeks before the interview, and Figure 1 shows the rates of the three diseases in the governorates of Egypt. In the following, we provide some more information about the three diseases, which are used as response variables and information about the covariates considered in the case study.

Diarrhea. Diarrheal disease, caused by the poor condition of water and sanitation, is a common public health problem in Third World countries. It is a variety of micro-organisms including viruses, bacteria, and protozoans that cause diarrhea, affecting people’s health through loss of water and electrolytes. This leads to dehydration and in disastrous preconditions also to death. In Egypt, the widespread use of oral rehydration therapy has successfully reduced the severity of diarrheal episodes and sharply reduced the number of subsequent deaths. However, overall diarrheal disease has not declined. In the 2003 EDHS, mothers were asked whether any of their children less than 5 years of age had had diarrhea at any time during the 2-week period before the survey.

Fever. Infection is the most common cause of fever in children. Most fevers in babies and children are caused by a viral (germ) infection. Common viral and bacterial illnesses like colds, gastroenteritis, ear infections, croup, and bronchitis are the most likely illnesses to cause fever.

Cough. Cough and breathing difficulties are common problems in young children. Recent literature indicates that breastfed children who had a cough or cold may have difficulties in feeding. Breastfeeding however, could help fight diseases. Along with diarrhea, acute respiratory infection (ARI), particularly pneumonia, is a common cause of death in infants and young children.

Categorical covariates. Table 2 provides information on categorical socioeconomic and bio-demographic covariates, their categories, frequencies, and the coding used in the regression models. Although the availability of radio, electricity, type of toilet, and drinking water are included in M1 and M2, we considered only the wealth index instead in the statistical analysis of M3 and LVM. Wealth index is measured as a poverty status in the DHS survey, rather than income or consumption. The wealth index is a composite measure of the cumulative living standard of the household. It contains a household’s ownership of selected assets, such as TV, radio, materials used for housing construction, and types of water access, and sanitation facilities.

The following continuous covariates have possibly non-linear effects on diseases.

Child’s age (chage). The age of a child has a significant effect on its morbidity as reported in many previous studies.
According to the World Health Organization (WHO) children should receive all recommended vaccines by 12 months of age. Figure 2A shows that many cases fall in the age group 10–12 months. The effect of child’s age is also affected by the weaning crisis where children are exposed to feeding other than breast milk.

**Mother’s body mass index (BMI).** Body mass index (BMI) varies with the woman’s age, and it is somewhat higher among urban women than among rural women. Studies show that this coexistence of under- and overnutrition exists not only at the societal but also the household level. The range of overweight mothers is remarkably large, even within a region. For instance, 55% of mothers are overweight in Egypt. Figure 2B shows that there are many overweight mothers (between 27–30 BMI) in the data set for Egypt. Motivated by the conceptual framework of Mosley and Chen, we consider mother’s BMI as a potential maternal factor, and we want to explore its effect on child morbidity.

**Mother’s age at birth (magb).** This is an important variable to fertility because it marks the onset of the childbearing process. In a typical Middle Eastern culture, magb is expected to be highly correlated with age at first marriage. Delay in magb may indicate late establishment of marriage and hence implies shortening of the reproductive period and consequential reduced fertility. Figure 2C shows the kernel density estimates of magb. It reflects the effect of the increasing age of the mother at birth; a few mothers fall in the age group (12–19). It indicates motherhood between ages 18–24, and also shows that only a small percentage of women older than age 25 had given birth at the time of the survey.

**Spatial covariates.** The information of the geographic location (governorate) where the ill child lives at the time of interview is a significant contribution of the DHS data set to understanding child disease in Egypt. Table 3 provides the regional variation of the three types of diseases in Egypt. This application includes 20 governorates. The Egyptian regions used in this study and in previous studies are metropolitan or urban governorates (Cairo, Alexandria, Suez, and Port Said), Lower Egypt, Upper Egypt, and bordering areas. Figure 1 shows that Lower Egypt, essentially some districts in Nile Delta, is associated with significantly higher rates of illness. Red areas indicate that

![Figure 2](image-url)
there is a negligible effect within these areas, green areas reflect a strong effect in these regions, and gray areas indicate that no children live in these regions, according to the data set.

### BAYESIAN GEOADDITIVE PROBIT AND LATENT VARIABLE MODELS

Geoadditive regression models extend (generalized) linear models for various types of response variables by adding non-parametric terms for nonlinear effects of continuous covariates and geographic effects of a spatial variable to the usual linear part of the predictor. Similarly, predictors in latent variable models can be extended to geoadditive predictors. In the following, we focus on probit models for binary responses, but in general the approach also covers models with continuous, ordered categorical and count variables as observed responses. This section provides a compact presentation of this model.

Let \( y_1, \ldots, y_p \) denote \( p \) observable binary responses, such as the three disease indicators in our case study, and \( x_{1j}, \ldots, x_{pj} \) corresponding covariate vectors. Separate probit models with conventional linear predictors can be defined through

\[
P(y_j = 1|x_j) = \Phi(\beta_{0j} + x_j' \beta_j), \quad j = 1, \ldots, p. \tag{1}
\]

where \( \Phi \) is the standard normal distribution function.

Separate geoadditive probit models are obtained by extending the linear predictor in (1) to a geoadditive predictor.

\[
\eta_{ji}^{geo} = \beta_{0j} + x_j'(\beta_j + f_1(z_{1j}) + \cdots + f_k(z_{kj}) + f_{geo}(s_j)). \tag{2}
\]

Thus, geoadditive probit models are defined through

\[
P(y_j|\eta_{ji}^{geo}) = \phi(\eta_{ji}^{geo})
\]

with the geoadditive predictor (2). The smooth functions \( f_1', \ldots, f_k' \) represent nonlinear effects of continuous covariates \( z_{1j}, \ldots, z_{kj} \). For simplicity, we only consider the case that these covariates are the same for each predictor \( \eta_{ji}^{geo}, j = 1, \ldots, p. \) The function \( f_{geo} \) represents the geographic effect of a spatial variable \( s \in [1, \ldots, d] \), indicating regions or districts in a country. The geographic effect can be interpreted as a surrogate for unobserved variables with geographic information, incomplete or not covered by observable covariates. It may be split into a structured part \( f_{geo} \), for correlated spatial effects, and an unstructured part \( f_{geo}^{un} \), for uncorrelated, local spatial effects. Such an approach has been applied successfully to diseases in developing countries. Details on priors for modeling the functions and the spatial effect are provided in the Appendix.

The unknown parameters and functions \( f_1', \ldots, f_k' \) and \( f_{geo} \) have to be estimated from the data. We follow a semi-parametric Bayesian approach as developed in Fahrmeir and Lang, and Brezger and Lang. We assume diffuse, non-informative priors based on Markov chain Monte Carlo (MCMC) techniques \( p(\beta_{0j}) \propto \text{const}, p(\beta_j) \propto \text{const} \). Functions \( f_1', \ldots, f_k' \) follow P-spline priors, and the geographic effect \( f_{geo} \) is modeled through a Markov random field. We use MCMC simulations to draw samples from the posterior. Statistical inference is done by means of MCMC techniques in a full Bayesian setting. Full Bayesian inference is based on the entire posterior distribution

\[
p(\beta, \tau^2, y | y) \propto p(y | \beta, \tau^2) p(\beta, \tau^2, \gamma), \tag{3}
\]

where \( \beta = (\beta_0, \ldots, \beta_j) \) and \( \tau^2 = \tau_1^2, \ldots, \tau_2^2 \) denote parameter vectors for function evaluations and variance; see also Khatab for more details.

### Latent variable models for binary responses

A drawback of separate probit models for each of the binary responses \( y_j \) introduced so far is that association among \( y_1, \ldots, y_p \) can only be captured by joint covariates. Latent variable models, as introduced in this section, automatically induce correlation among the responses.

The basic idea of factor analysis and LVM is that the vector of the \( p \) observable variables can be represented, at least partly, by one or more latent factors or variables \( \psi \) with a lower dimension. As in our case study, where we introduce the latent variable \( \psi \) “health status” we only consider a one-dimensional latent variable for simplicity. Extension to multi-dimensional latent variables and models with different types of observable responses are presented in Fahrmeir and Raach and in Fahrmeir and Khatab.

The LVMs in our study extend the separate probit model (1) by adding the effects of a common latent variable \( \psi \) to the linear predictor, resulting in the observation models

\[
p(y_j = 1|x_j, \psi) = \phi(\beta_{0j} + x_j' \beta_j + \lambda_j \psi). \tag{4}
\]

The covariates \( x_j \) have direct effects \( \beta_j \) on the responses \( y_j \), similar as in separate probit models. The common latent variable \( \psi \) automatically induces correlation among the responses. The effects \( \lambda_j \) of \( \psi \) are usually called factor loadings. The observation is supplement through a structural model for the latent variable. Geoadditive latent variable probit models assume a geoadditive structural model

\[
\psi = u'\alpha + f(s) + \cdots + f(w_i) + f_{geo}(s) + \delta, \tag{5}
\]

with iid. Gaussian errors \( \delta \sim N(0,1) \). For identifiability reasons it is assumed that \( \text{var}(\delta) = 1 \), and that the predictor for \( \psi \) contains no intercept term. The additional covariates \( u, w_1, \ldots, w_k \) and the location variable \( s \) act directly on the latent variable \( \psi \), but indirectly on the observable responses. The nonlinear functions and the spatial effect are modeled through similar priors.
as for separate geoadditive probit models, see the Appendix. The LVMs of the form (4) and (5), restricted to a linear structural model \( v = \alpha' \alpha \), have been suggested previously by Sammel et al.

They also provide motivation and some guidance, which covariates might be kept in the measurement model (4) and which covariates should be relegated to the structural model. Currently, there are no automated purely data driven tools for model checking and diagnostics available for deciding on this. Thus, the decision is based on substantive reasoning in combination with more informal statistical arguments. From a pragmatic point of view it would be desirable to relegate as many covariates to the structural model as possible: This leads to more parsimonious models with less parameters and will allow explaining the association between and variability of indicators \( y_i \) through common effects acting via the latent variables. Our current strategy is as follows: We first fit separate probit models, as in the case study of this work. Then, we relegate covariates with similar effects or patterns to the structural model while the rest is kept in the measurement model. Statistical methodological development for formal model checking and choice in complex hierarchical models, in particular in LVMs, is desirable but just at the beginning.

Statistical analyses and results. Statistical analyses were performed in two steps:

First, we fitted separate geoadditive probit models to the following three diseases: diarrhea, fever, and cough. A main purpose of this step was model selection, to model effects of the continuous covariates, and to see if there are sizeable spatial effects. Based on preliminary exploratory analyses not shown we used the deviance information criterion (DIC) of Spiegelhalter and others to select models in a formal way. As a first step, we apply separate geoadditive probit models for diarrhea, cough, and fever. The results are shown in Tables 5–7. In the second step, we then applied geoadditive probit LVMs to analyze the data. Although the DIC is now commonly accepted as a standard tool for selecting probit or logit models, its performance for LVM model choice is not yet well understood. We therefore proceeded as described at the end of the Bayesian Geoadditive Probit and Latent Variable Models section: If the effects of covariates turned out to be significantly different (in terms of confidence intervals) for the three diseases, we decided to keep them in the measurement model, otherwise covariates were included in the geoadditive predictor of the structural equation for the latent variable. All nonlinear effects and the spatial effect are included in the structural model.

Analyses with separate geoadditive models. We present results for the following probit models, selected from a longer hierarchy of models. The responses \( y_{ij} \), \( j = 1 \) (diarrhea), \( 2 \) (fever), \( 3 \) (cough) are coded as

\[
y_{ij} = \begin{cases} 
1 & \text{if child had disease } j \text{ two weeks prior to the survey} \\
0 & \text{if not} 
\end{cases} 
\]  

(6)

The following covariates were considered in the analysis:

Metrical covariates

- \textit{Chage}: Child’s age in months.
- \textit{BMI}: Mother’s body mass index.
- \textit{Mageb}: Mother’s age at birth.
- \textit{trepr}: Whether mother had treatment during pregnancy: yes or no (reference category).
- \textit{anvis}: Whether mother had antenatal care: yes or no (reference category).
- \textit{urban}: Locality where respondent lives: urban or rural (reference category).
- \textit{wealth index}: Poverty status is measured in the DHS survey in terms of a wealth index \( j \) first quintile (reference category) to the 5th quintile.
- \textit{Feed type}: Type of breastfeeding: no breastfeeding (reference category), mixed feeding, and exclusive breastfeeding.
- \textit{Work}: Mother’s current working status: working or not (reference category).

Spatial covariate

\textit{reg}: Governorate where respondent resides.

The predictors of the models considered in this section are as follows:

- \( M_0 \): Includes only district-specific effects.
  
  \[ M_0: \eta_{ij} = \beta_{0i} + f_{we} (\text{reg}) + f_{wean} (\text{reg}) \]  
  (7)

- \( M_1 \): Includes all categorical covariates and the metrical covariates.
  
  \[ M_1: \eta_{ij} = \beta_{0i} + f_j (\text{Chage}) + f_i (\text{BMI}) + f_j (\text{Mageb}) + w_{ij} \gamma_i \]  
  (8)

- \( M_2 \): Adds district-specific effects to Model 1.
  
  \[ M_2: \eta_{ij} = \beta_{0i} + f_j (\text{Chage}) + f_i (\text{BMI}) + f_j (\text{Mageb}) + f_{we} (\text{reg}) + f_{wean} (\text{reg}) + w_{ij} \gamma_i \]  
  (9)

- \( M_3 \): Includes all the bio-demographic and health factors. In model M3 the vector \( w \) is reduced to the vector \( z \) by omitting factors of education, type of toilet, availability of radio, availability of electricity, and source of water, and include instead the wealth index and the type of feed. The metrical covariates child’s age, mother’s BMI, and mother’s age at birth are allowed to have a nonlinear effect on the diseases of the child as well as the spatial effects \( f_{we} \) and \( f_{wean} \).

In these models the covariate vector \( w \) in models M1 and M2 contains all the bio-demographic and health factors. In model M3 the vector \( w \) is reduced to the vector \( z \) by omitting factors of education, type of toilet, availability of radio, availability of electricity, and source of water, and include instead the wealth index and the type of feed. The metrical covariates child’s age, mother’s BMI, and mother’s age at birth are allowed to have a nonlinear effect on the diseases of the child as well as the spatial effects \( f_{we} \) and \( f_{wean} \).

It turned out that model M3 for each type of disease is superior in terms of the DIC.

Sensitivity analysis. It is known that the Markov Random Field prior for spatial covariates works well if there are many neighbors for the spatial units. However, this is not the case for Egypt, where there are few governorates and neighbors. Therefore, we carried out a sensitivity analysis for the choice of the prior for the spatial effects. It turned out that the results of the spatial effects remained stable for the separate models and also for the latent variable models.

Results. In the preliminary analysis, we aim to separate the two kinds of spatial effects included in model M0 to estimate a structured and an unstructured effect. In a further step, we include the categorical covariates and the metrical covariates in the analysis as in M1, M2, and M3. The results for these models are given in Khatab. We focus on the results of M3, which is the best model in terms of DIC, see Table 4. The
results for the categorical covariates are shown in Tables 5–7 for the three diseases, respectively.

Figures 3–5 display the nonlinear effects of the continuous covariates on the three diseases, respectively, and Figures 6–8 show district variation in prevalence of diarrhea, fever, and cough using geoadditive separate analyses.

**Diarrhea.** Table 5 displays the estimated effects of the categorical variables (male, urban, mother’s working status, mother’s treatment during pregnancy, antenatal visits, mother’s education, wealth index, and the type of feeding) on diarrheal disease in Egypt. The fixed effects show that risk of diarrhea is lower among female infants than among male infants. The results indicate that treatment of mother during pregnancy and antenatal visits significantly increase the risk for diarrhea. Later, we will discuss this seemingly counter-intuitive effect and antenatal visits significantly increase the risk for diarrhea. The upper quintile category of wealth index and exclusive breastfeeding are associated with a lower risk of diarrhea morbidity. The analysis also suggests that working status of mother has little or no significant effect.

With regard to the nonlinear effects, Figure 3A–C show the (nonlinear) effects of age of the child, mother’s body mass index, and the mother’s age at birth for model M3, respectively, modeled through Bayesian P-splines. The nonlinear effect of a child’s age suggests that there is continuous and serious worsening of children’s health status up to about 11 months of age, with an almost linear decline thereafter. The impact of a mother’s BMI on diarrhea is only slight. There is some evidence that children of mothers with a BMI less than 25 face a lower risk of disease (even though there are few mothers with BMI between 15 and 20). For BMI that are larger than 43–45, there are few observations and the credible intervals get wider. A somewhat higher risk for diarrhea seems to exist for mothers with a BMI between 27 and 30, where a bump appears. In addition, we find the influence of mother’s age (Figure 3C) on diarrhea in Egypt seems to be in the form of an inverse U-shape. It shows that the mother’s age has a slight impact on diarrhea; however, the children from mothers who are in the age group (18–22 years) are at a higher risk of diarrhea compared with children from mothers in other age groups.

With regard to spatial effects, Figure 6 displays the estimates of the spatial effect (the levels correspond to “high risk of morbidity” (green colored) and “low risk” (red colored)) for Egypt. The colored maps show posterior means of structured random effects on diarrhea (Figure 6B) and its corresponding posterior mean of unstructured random effects (Figure 6A). For the model M3 for the diarrhea disease, the geographic pattern of regions in Figure 6A reflects the estimated posterior means of the structured random effects on diarrhea. Obviously, there exists a district-specific geographic variation in the level of the disease in Egypt (Figure 6) based on the 2003 EDHS. The pattern reveals that significant high rates of illness are associated with the Upper Egypt area (Minya, Amarna, Luxor, Esna, Edfu, Aswan, etc.), some cities and rural areas in the Nile Delta and in Eastern Cairo (Sinai). Upper Egypt implies a relative higher risk of having a diarrhea disease and knowing the characteristics of the region, the result is not surprising. The left panel also reveals a higher risk of diarrhea morbidity

---

**Table 4**  
The deviance information criterion (DIC)  
<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>pD</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>6364.46</td>
<td>15.45</td>
<td>6395.38</td>
</tr>
<tr>
<td>M1</td>
<td>5433.27</td>
<td>36.53</td>
<td>5506.34</td>
</tr>
<tr>
<td>M2</td>
<td>5432.74</td>
<td>36.91</td>
<td>5506.55</td>
</tr>
<tr>
<td>M3</td>
<td>5308.83</td>
<td>45.30</td>
<td>5401.80</td>
</tr>
</tbody>
</table>

**Table 5**  
Fixed effects of model M3 on diarrhea-Egypt  
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>Median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const</td>
<td>−1.92*</td>
<td>0.15</td>
<td>−2.26</td>
<td>−1.91</td>
<td>−1.65</td>
</tr>
<tr>
<td>Urban</td>
<td>−0.063</td>
<td>0.068</td>
<td>−0.188</td>
<td>−0.064</td>
<td>0.075</td>
</tr>
<tr>
<td>Male</td>
<td>0.112*</td>
<td>0.035</td>
<td>0.040</td>
<td>0.113</td>
<td>0.178</td>
</tr>
<tr>
<td>Work</td>
<td>−0.009</td>
<td>0.045</td>
<td>−0.100</td>
<td>−0.005</td>
<td>0.074</td>
</tr>
<tr>
<td>Anvis</td>
<td>0.148*</td>
<td>0.037</td>
<td>0.074</td>
<td>0.149</td>
<td>0.219</td>
</tr>
<tr>
<td>Trepr</td>
<td>0.065*</td>
<td>0.031</td>
<td>0.002</td>
<td>0.065</td>
<td>0.129</td>
</tr>
<tr>
<td>Mixedfeed</td>
<td>−0.026</td>
<td>0.057</td>
<td>−0.142</td>
<td>−0.030</td>
<td>0.094</td>
</tr>
<tr>
<td>Exclusivefeed</td>
<td>−0.253*</td>
<td>0.086</td>
<td>−0.42</td>
<td>−0.253</td>
<td>−0.072</td>
</tr>
<tr>
<td>Wealthindex2</td>
<td>0.140</td>
<td>0.072</td>
<td>0.083</td>
<td>0.144</td>
<td>0.277</td>
</tr>
<tr>
<td>Wealthindex3</td>
<td>−0.018</td>
<td>0.075</td>
<td>−0.153</td>
<td>−0.018</td>
<td>0.132</td>
</tr>
<tr>
<td>Wealthindex4</td>
<td>−0.127</td>
<td>0.076</td>
<td>−0.269</td>
<td>−0.129</td>
<td>0.034</td>
</tr>
<tr>
<td>Wealthindex5</td>
<td>−0.339*</td>
<td>0.105</td>
<td>−0.538</td>
<td>−0.340</td>
<td>−0.143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>Median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const</td>
<td>−0.828*</td>
<td>0.085</td>
<td>−1.005</td>
<td>−0.829</td>
<td>−0.651</td>
</tr>
<tr>
<td>Urban</td>
<td>0.046</td>
<td>0.047</td>
<td>−0.047</td>
<td>0.045</td>
<td>0.134</td>
</tr>
<tr>
<td>Male</td>
<td>0.067*</td>
<td>0.029</td>
<td>0.012</td>
<td>0.064</td>
<td>0.129</td>
</tr>
<tr>
<td>Work</td>
<td>0.042</td>
<td>0.033</td>
<td>−0.020</td>
<td>0.043</td>
<td>0.103</td>
</tr>
<tr>
<td>Anvis</td>
<td>0.151*</td>
<td>0.039</td>
<td>0.073</td>
<td>0.151</td>
<td>0.232</td>
</tr>
<tr>
<td>Trepr</td>
<td>0.024*</td>
<td>0.030</td>
<td>−0.033</td>
<td>0.027</td>
<td>0.087</td>
</tr>
<tr>
<td>Mixedfeed</td>
<td>−0.053</td>
<td>0.05</td>
<td>−0.157</td>
<td>−0.052</td>
<td>0.050</td>
</tr>
<tr>
<td>Exclusivefeed</td>
<td>−0.194*</td>
<td>0.069</td>
<td>−0.322</td>
<td>−0.195</td>
<td>−0.063</td>
</tr>
<tr>
<td>Wealthindex2</td>
<td>0.140</td>
<td>0.078</td>
<td>−0.030</td>
<td>0.144</td>
<td>0.291</td>
</tr>
<tr>
<td>Wealthindex3</td>
<td>−0.018</td>
<td>0.075</td>
<td>−0.153</td>
<td>−0.018</td>
<td>0.132</td>
</tr>
<tr>
<td>Wealthindex4</td>
<td>−0.122*</td>
<td>0.061</td>
<td>−0.245</td>
<td>−0.122</td>
<td>−0.002</td>
</tr>
<tr>
<td>Wealthindex5</td>
<td>−0.165*</td>
<td>0.075</td>
<td>−0.306584</td>
<td>−0.166</td>
<td>−0.022</td>
</tr>
</tbody>
</table>

* denotes that effects are significant at a 95% significance level.  
* Statistically significant at 0.05%.
in the upper area despite being surrounded by some districts with lower risk.

**Fever.** The fixed parameters show that the prevalence of fever in Egypt (Table 6) is higher among male infants and among children from mothers with antenatal visits during pregnancy. The upper quintile category of wealth index in the household, representing families in the group of highest household economic status, is associated with a lower risk of fever morbidity. However, the results suggest that whether the children are from mothers who are working or not, and rural or urban residence have only a slight influence on fever morbidity in Egypt. If mothers fed their child with breast milk only (exclusive feed), the risk of childhood fever is significantly decreased. In addition, antenatal care during pregnancy has a positive significant effect on fever.

Figure 4 shows the nonlinear effects of a child’s age on fever. The impact of a child’s age is quite similar for the three models. It shows that deterioration sets in right after birth and continues up to 11–12 months, but then the age effect declines more or less steadily until 25–26 months. Mother’s BMI only has a slight significant impact on child health status. It declines for mothers with a BMI of less than 20, and is less pronounced for mothers with BMI between 20 and 35, despite a bump between a BMI of 30 and 35, which is caused by overweight mothers. For a BMI of 40 and over there are only a few observations, reflected in wide credible intervals. Unexpectedly, the effect of mother’s BMI in the three models turns out to be almost linear.

Concerning the nonlinear effect of mother’s age at birth on fever, Figure 4 displays that children from younger mothers (< 20 years of age) are at considerably higher risk than children from mothers from the middle-aged group (25–35).

The geographic pattern of district-specific effects for fever in Figure 7 indicates that significant high illness rates are associated with the Egyptian governorates Suez, El Arish, Ismalia, and Sinia “in the southwestern area.” There is a variation in the level of illness rates of children in Egypt. This variation could be attributed to environmental risks, which in turn influence exposure to disease. The unstructured effects are similar to the structured effects. The gray area in these sparsely populated regions indicates that no children are included in the EDHS sample.
Cough. The results indicate that children from mothers who attended antenatal care during pregnancy, living in an urban area, and currently working, face a high rate of cough disease compared with children from mothers who are not working, living in a rural area, and not attending any care. The results also suggest that the upper quintile category of wealth index and exclusive feed reduce the risk of cough disease in Egypt. Boys less than 5 years of age are more susceptible to cough than girls. The rest of categorical covariates have either a negligible or an insignificant effect on cough.

The nonlinear effect of child’s age for model M3 has a similar pattern as for diarrhea and fever. The same is true for mother’s BMI and mother’s age at birth.

The spatial effect on cough in Figure 8 suggests that significantly high rates of cough illness are associated with Damietta, Dakhalia, and Esmaliyia.

DISCUSSION

Fixed effects. As for child’s gender, the probability of disease is believed to be higher for males, because of biologic reasons. Furthermore, boys noticeably more so than girls are provided with treatment. For some countries studies show higher female mortality, indicating gender discrimination. For Egypt, the results show that the risk for the three types of diseases diarrhea, fever, and cough is more widespread in male children. Similar findings are also reported in previous studies.

The effects of urban versus rural place of residence are different for the three diseases: For diarrhea and fever, living in urban areas lowers the risk. For cough, the effect is significant for children from urban versus rural areas. These results support the important role of public health policy in rural-urban disparities.

Mothers who attend a clinic to receive antenatal care during the period of pregnancy are expected to have lower problems than those who do not receive any care. The results for Egypt, however, suggest the opposite: the factor antenatal visit has a positive effect on the indicators of disease!

There are several possible reasons or hypotheses to explain this counter-intuitive effect. A possible reason could be that only a few mothers had frequent antenatal visits during their pregnancy. In addition, rural Lower Egypt had the lowest number of regular check-ups and the highest percentage of medical problems, as reported in previous studies. Another possible hypothesis was suggested: First, antenatal visits could be a sign of problem pregnancies, making subsequent childhood illnesses more likely. Second, there might be a perception bias of morbidity: Mothers who watch their children’s illness more closely are more likely to take them to antenatal visits and, on the other hand, may report them ill in interviews, although their children’s illness might not be so serious.

Figure 5. Nonlinear effects of child’s age, mother’s BMI, and mother’s age for M3 on cough for Egypt using probit model.

Figure 6. Maps of Egypt for diarrhea showing (A) unstructured and (B) structured spatial effects for M3 using probit model. This figure appears in color at www.ajtmh.org.
In contrast, mothers who observe their children’s illness less closely are less likely to have antenatal visits and may not report them ill, even if they are. Such perception effects might vary over income groups. To study the latter hypothesis, we included a wealth index as a further covariate. The counter-intuitive effect remains, however.

Concerning current working status of mothers, the results suggest a significant effect of this variable on cough morbidity in Egypt. However, the effect is positive. The problem lies where mothers engage in out-of-home employment. This curtails the duration of full breastfeeding and necessitates recently introduced supplementary feeding, often by the illiterate caretakers, and that could have a side effect on the health of the child in the early months.

Nonlinear effects. In general, the results show that the risk of having diseases in the 2-week reference period reaches its peak at 11 months of age and then begins to fall with increasing age of the child. This pattern resembles those found in many studies of sub-Saharan Africa. The prevalence of disease was found to be highest among children 6–12 months of age, the period when most children are weaned. In addition to breast milk, inborn immunity and less exposure to contaminated agents during the early period also contribute to the lower prevalence of diarrhea. On the other hand, prevalence is quite high when the child has lost inborn immunity and when the child is exposed to different types of infections by eating food prepared with contaminated water and from an unhealthy environment.

Likewise, the effect of mother’s age at birth is almost linear in Egypt, particularly in the interval age between 20 and 27 years. The curve has a slight bathtub shape, indicating that children from younger mothers (12–20) have higher risk, compared with mothers 20–35 years of age. The results reflect a slight effect of mother’s age at birth on the morbidity of children.

In the literature, the influence of the BMI of the mother is sometimes expected to be inversely U-shaped. Parents with low BMI values are malnourished and are therefore likely to have undernourished and weak children. At the same time, very high BMI values indicate poor quality of the food and hence, may also imply weakness of the children in our study. The results of Egypt indicate that a mother’s BMI of 27–30 greatly increases the effect on child morbidity. Beyond a BMI of 30, the effect remains stable at a low level. The higher impact of BMI through the interval between 27 and 30, indicates poor quality of food for mothers and hence, may imply malnutrition of the child and affect the health of the child.

Spatial effects. The Egyptian regions used in this study and in previous studies are metropolitan, Lower Egypt, Upper Egypt, and border areas. Ninety-five percent of the population of Egypt lives in the first three regions. The metropolitan governorates essentially comprise the four major cities of Cairo, Alexandria, Port-Said, and Suez, all in northern Egypt. Lower Egypt (essentially the region of the Nile Delta) is also in the northern part of Egypt, and Upper Egypt is the area south of Cairo, with governorates largely following the meandering upper parts of the Nile. The border areas are the less populated desert areas bordering the Red Sea, the Sinai, and the vast Marsa Matruh and El Wadi El Gadid areas west of the Nile. Generally, childhood diseases appear to have

Figure 7. Maps of Egypt for fever showing (A) unstructured and (B) structured spatial effects for M3 using probit model. This figure appears in color at www.ajtmh.org.

Figure 8. Maps of Egypt for cough showing (A) unstructured and (B) structured spatial effects for M3 using probit model. This figure appears in color at www.ajtmh.org.
a higher influence in the north-eastern region, affecting the most of districts there.

The level of health care is considered to be low in these areas. In particular, supply of water is available only to 67% of the residents compared with 86–99% in urban governorates and in Lower Egypt.

However, most of the poor were found in Upper rural Egypt with the highest rate of illnesses. The 5.5 million poor people, out of 10.7 million in all, live in these regions, whereas 1.4 million of the poor live in the urban parts of Upper Egypt. Moreover, the report indicates that about 17% of the Egyptian population was poor in the year 2000. The lower standard of living in these areas has a direct impact on the rate of illiteracy and on the educational level of mothers, leading to more poverty in these areas and a lower level of sanitation and public health.

Analyses with latent variable models. As previously discussed, we now investigate how the three diseases can be interpreted as indicators of a latent variable \( u \) “health status” of children, how much of the variation of \( u \) can be explained through a geoadditive predictor, and which covariates have a direct effect on the disease indicators. This concept does not only allow us to analyze the impact of covariates on health status, it also automatically introduces correlation among disease indicators. To demonstrate the latter property, we first consider a classic factor analytic model without any covariates.

\[ p(y_{ij} = 1 | u_i) = \Phi(\lambda_j u_i), \quad u_i \sim N(0, 1) \]  

\[ \text{Table 8} \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever ( \lambda_{11} )</td>
<td>2.2</td>
<td>0.34</td>
<td>1.78</td>
<td>3.03</td>
</tr>
<tr>
<td>Cough ( \lambda_{21} )</td>
<td>0.87</td>
<td>0.04</td>
<td>0.77</td>
<td>0.959</td>
</tr>
<tr>
<td>Diarrhea ( \lambda_{31} )</td>
<td>0.67</td>
<td>0.03</td>
<td>0.616</td>
<td>0.72</td>
</tr>
</tbody>
</table>

* Statistically significant at 0.05%.

\[ \text{Table 9} \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever ( \lambda_{11} )</td>
<td>1.34*</td>
<td>0.093</td>
<td>1.12</td>
<td>1.48</td>
</tr>
<tr>
<td>Cough ( \lambda_{22} )</td>
<td>0.9*</td>
<td>0.04</td>
<td>0.71</td>
<td>0.91</td>
</tr>
<tr>
<td>Diarrhea ( \lambda_{32} )</td>
<td>0.70*</td>
<td>0.04</td>
<td>0.71</td>
<td>0.89</td>
</tr>
</tbody>
</table>

* Statistically significant at 0.05%.

\[ \text{Table 10} \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever ( \lambda_{11} )</td>
<td>1.273*</td>
<td>0.099</td>
<td>1.090</td>
<td>1.487</td>
</tr>
<tr>
<td>Cough ( \lambda_{22} )</td>
<td>0.824*</td>
<td>0.040</td>
<td>0.746</td>
<td>0.911</td>
</tr>
<tr>
<td>Diarrhea ( \lambda_{32} )</td>
<td>0.796*</td>
<td>0.047</td>
<td>0.706</td>
<td>0.889</td>
</tr>
</tbody>
</table>

* Statistically significant at 0.05%.

Figure 9. Nonlinear effects (from top to bottom): child’s age, mother’s BMI, and mother’s age at birth for model LVM1; child’s age, mother’s BMI, and mother’s age at birth for model LVM3 on the indicators of a latent variable “health status” of children disease for Egypt using Bayesian latent variable model for binary responses.
Latent variable models offer a new methodology for considering special types of diseases as indicators for latent morbidity and to flexibly model covariate and spatial effects on this latent variable. Compared with separate geadditive models for the indicator variables, latent variable models can be advantageous from a statistical and a substantive perspective. Common latent variables automatically induce association between indicator variables not covered by covariates. Furthermore, they are more parsimonious, i.e., they require less parameters and unknown functions. Concerning interpretation and conclusions, advantages perhaps can be best explained by considering the final model LVM3: They allow public health policy to focus on the most important (indirect) covariates that have a general impact on morbidity. The effects of the direct covariate “urban” showed that additional but different public health policies are required depending on the type of disease. In Egypt, therefore, we excluded these covariates in model LVM3 (Table 10).

The pattern for nonlinear effects on the latent variable health status closely resembles the patterns of separate analyses. Furthermore, there is no noticeable difference between the nonlinear effects by model LVM2 and model LVM3 see Figure 9. Therefore, only the results of model LVM3 are reported here.

The spatial effect is displayed in Figure 10, and shows that the northeast has an influence on the latent variable associated with high illness rates. These areas face problems with health conditions, level of sanitation, and water supplies that could lead to a high level of infections among children less than 5 years of age living in these areas.

CONCLUSION

In this work, we studied socioeconomic, public health, and spatial determinants of morbidity measured through prevalence of fever, cough, and diarrhea in Egypt. Our analyses show that geadditive models are needed to adequately assess nonlinear covariate effects and geographic effects within a joint model. With a traditional regression model these effects are difficult if not impossible to model and to detect.
diarrhea, and cough obtained retrospectively from cross-sectional studies.

First, seasonal differences of occurrence in diarrhea cannot to be taken into account in such studies. Longitudinal studies may be more appropriate to provide data in different seasons. Second, during the survey, neither the children were examined nor mothers were given a precise definition of what constitutes an episode of various diseases. In addition, we have no sufficient information about the children who died before the survey, and whether the cause of dying was the kind of diseases that are reported here or not. The questions measure (in the DHS) the mother’s perception of her child’s health rather than morbidity according to clinical examination. This may create variations among different socioeconomic groups because perception of illness is not the same across different social groups. Third, loss of memory of events and misinterpretation of the reference period can also contribute to the problems associated with the prevalence of diarrhea.36,27

Lessons learned and further research. The prevalence of diseases was found to be highest among children 11 or 12 months of age, the period when most children weaned. Therefore, it is recommended that mothers exclusively breastfeed their babies for the first 2 years of life. In developing countries, continued breastfeeding is recommended up to at least 2 years of age with the timely addition of appropriate complementary food at 6 months of age.

- Children of mothers who are working are more likely to have health problems. Therefore, mothers should give their children more time and more care during the first years of life.
- The children of younger mothers have higher risk, compared with mothers 22–35 years of age.
- The higher impact of BMI through the interval between 27 and 30 indicates poor quality of food for mothers.
- Urbanized children are at a lower risk of diarrhea, but they are at a higher risk of cough morbidity.
- The spatial effect suggests the need to give more attention to some areas that have high rates of diseases, such as the Nile Delta, Upper Egypt, and southeastern in Egypt. These areas are more likely to have a higher proportion of morbidity compared with other areas, caused by poor health facilities and complications during childhood or even careless and misdiagnosis during hospital care. Therefore, the most important issues to address in these areas are health care, proper food, and raising the educational level of parents. Governments should improve socioeconomic conditions. Because, if living standards are improved, there will be better health care and a reduction in infant and, child diseases, child malnutrition, and child mortality.

As a further work, we will study the geographic and socioeconomic determinants of childhood malnutrition in Egypt, also using the latent variable models. Furthermore, it seems useful to extend the latent variable model by using two factors instead of one to take the correlations into account between health status and nutritional status using latent variable models with mixed binary and continuous indicators.

Received April 8, 2008. Accepted for publication February 19, 2009.

Acknowledgments: We thank the referees for helpful comments and suggestions, which helped to considerably improve the first version of the article. Our gratitude also goes to Ngianga-Bakwin Kandala for stimulating discussions on application of geoadditive models to malnutrition and morbidity.

Financial support: Part of this research was financially supported by the Munich Center of Health Sciences.

Authors’ addresses: Khaled Khatab, Pauwels Street 30, 52074 Aachen, Germany, E-mails: kkhatab@ukaachen.de and khaledkhatab314@yahoo.com, Tel: +49 241 80 36733, Fax: +49 241 8082587. Ludwig Fahrmeir, 33 Ludwig Street-80539 Munich, Germany, E-mail: Ludwig.Fahrmeir@stat.uni-muenchen.de, Tel: ++49 21806253, Fax: ++49 21805040.

REFERENCES

5. Demographic and Health Survey (EDHS) for Egypt, 2003. MEASURE DHS (Demographic and Health Surveys). Calverton, MD.


