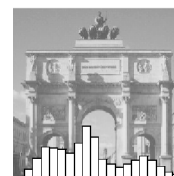




LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK



Khaled Khatab & Ludwig Fahrmeir

Analysis of Childhood Morbidity with Geoadditive Probit and Latent Variable Model: A case study for Egypt.

Technical Report Number 021, 2008
Department of Statistics
University of Munich

<http://www.stat.uni-muenchen.de>



Analysis of Childhood Morbidity with Geoadditive Probit and Latent Variable Model: A case study for Egypt.

Khaled Khatab and Ludwig Fahrmeir

**Department of Statistics. University of Munich, Ludwigstr. 33,
D-80539 Munich-Germany**

Abstract

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 provided 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 which affected the health of child in early years, various bio-demographic and socio-economic variables. In this paper we investigate the impact of such factors on childhood disease with flexible geoadditive models. These models allow 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 modelling and inference. As a first step we employ separate geoadditive probit models to the binary target variables for diarrhea, cough and fever using covariate information from the 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 modelling approach allows 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.

Keyword: Childhood diseases, developing countries, geoadditive regression model, latent variable models, MCMC.

1 Introduction

Childhood disease is among the most serious health issues facing developing countries, and it has an impact on future development. The main objective of this paper is to examine the impact of the socio-economic and bio-demographic factors on childhood disease, including geographical 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 Found Activities (UNPFA) and the U.S.Agency International Development (USAID). One of the main objectives of EDHS is to provide an up to date information on childhood disease. This intends to assist policy makers and administrators in evaluating and designing programs and improve 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 socio-economic, public health and geographical variables on disease of young children in developing countries with Egypt as a case study. Statistical analysis will be based on modern Bayesian approaches which allow to flexibly formulate 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 (2001) and Brezger and Lang (2005). In a second step, we use geoadditive latent variable models, recently suggested in Raach (2005) and Fahrmeir and Raach (2007). 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 modelling approach 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. Compared to previous results, our approach can provide additional and new insight of childhood morbidity and mortality in developing countries in general and, more specifically, in Egypt.

Previous studies on child disease have focused on various-socio-economic, demographic or health factors available in specific data sets, however most of these studies have neglected aspects for the spatial effects, see for instance see for instance Miller and Hirschhorn (1995), and Miller et al. (1994). Previous work on child disease in Egypt is restricted to few selected or specific towns and governorates. However, a notable exception is the study for Nigeria in Kandala et al. (2007), using separate geoadditive probit models for cough, fever and diarrhea. For such work, see Langsten and Hill (1994). Our case study, different from these previous works with respect to the following aspects: Firstly, the analysis studies spatial differentials of child disease at a highly disaggregated governorates level using a Bayesian approach for geoadditive models. This allows to incorporate covariates effects in a flexible semi-parametric way, which is not possible through the usual parametric approaches considered previous works. Secondly, a latent variable model (LVM) for health status based on binary disease indicators permits modelling of covariates effects on the latent variables through a flexible geoadditive predictor. All computations have been carried out with BayesX- Version 1.40 (Brezger et al., Lang 2005), STATA and R Programs using MCMCpackage see Raach.(2005) and Fahrmeir and Raach (2007). The rest of the paper is organized as follows: In Section 2 we discuss the data set and methods, while Section 3 describes geoadditive models and latent variable models. Section 4 contains data analysis and results for child disease in Egypt. Concluding remarks are given in Section 5.

2 Data

This paper 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, and 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 6661 children in Egypt. Each record consists of disease information and the list of covariates which could affected child's health. The EDHS data only permit to attribute child morbidity to specific disease in the last two weeks before the surveys. Table 2 shows an overview of the three common diseases in Egypt in the last two 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, as well as about the covariates considered in the case study.

Diarrhea

There is a variety of micro-organisms that could be the main cause of the diarrhea disease including viruses, bacteria and protozoans. Diarrhea affects the health of persons and causes to loss of water and electrolytes as well which are a leading cause of both dehydration and death in some other cases. It is a most public health problem which related to water and sanitation. In the 2003 EDHS, mothers were asked whether any of their children under five years of age had diarrhea at any time during the two-week period prior to the survey.

Fever

Most fevers in babies and children are caused by a viral (germ) infection. However, fever is less common and high fevers are unusual in young infants, and any fever should be considered a danger sign of very severe disease. The causes of fever could be an infection caused by germs like virus, parasites, or bacteria and vaccinations or immunization shots. Sometimes children have fever for no known reason.

Cough

Cough and difficult breathing are common problems in young children. The recent literature indicates that breastfed children who had a cough or cold may have difficulties in feeding, however breastfeeding could help to fight the diseases. Along with diarrhea, acute respiratory infection (ARI), particularly pneumonia, is a common cause of death among infants and young children (EDHS 2003).

Categorical covariates

Table 2 provides information on categorical socio-economic and bio-demographic covariates, their categories, frequencies and the coding used in the regression models.

Variable	Obs	Mean	Std. Dev.	0:had no diseases	1:had diseases
Diarrhea	6348	0.210	0.407	5.013(78.97)	1.335(21.03)
Fever	6348	0.323	0.467	4.297(67.69)	2.051(32.31)
Cough	6348	0.255	0.4361	4.725(74.43)	1.623(25.57)

Table 1: Overview of diseases in Egypt

Factor	n (%)	coding
Place of residence		
Urban	2237(33.58%)	1
Rural	4424(66.42%)	-1.ref
Child's sex		
Male	3487(52.35%)	1
Female	3174(47.65%)	-1.ref
Working		
Yes	1209(18.15%)	1
No	5452(81.85%)	-1.ref
Mother's education		
No,		
Incomp.prim,		
Comp.prim,		
Incomp.sec	4194(62.97%)	1
Compl.sec,		
Higher	2467(37.04%)	-1.ref
Pregnancy's treatment		
Yes	697(10.46%)	1
No	5964(89.54%)	-1.ref
Receive vaccination		
Yes	1737(25%)	1
No	56(0.8%)	-1.ref
Missing	75%	
Drinking water		
Controlled	5374(80.68%)	1
Not controlled	1287(19.32%)	-1.ref
Missing	1%	
Has radio		
Yes	5374(80.68%)	1
No	1559(19.32%)	-1.ref
Has electricity		
Yes	6203(93.12%)	1
No	458(6.88%)	-1.ref
Toilet facility		
Own flush toile	1768(28%)	1
No toilet	4511(71.8%)	-1.ref
Missing	1%	
Antenatal visit		
Yes	4181(63%)	1
No	2342(35%)	-1.ref
Missing	2%	

Table 2: Factors analyzed in diseases study



Figure 1: Maps of Egypt showing the rates of diarrhea disease (top left), rate of Fever disease (top right) and rate of cough disease(bottom)

The following continuous covariates have possibly nonlinear effects on diseases.

Child's age (chage)

The prevalence of diseases depends on age. According to the world health organization (WHO), children should receive the complete schedule of recommended vaccinations by 12 months of age. Figure 2 (left panel) shows that there are many cases fall in the age group 10-12 months.

Mother's body mass index (BMI)

Body mass index varies with the woman's age, and it is somewhat higher among urban women than among rural women (EDHS 2003). Studies show that this coexistence of under- and overnutrition exists not only at the societal but also

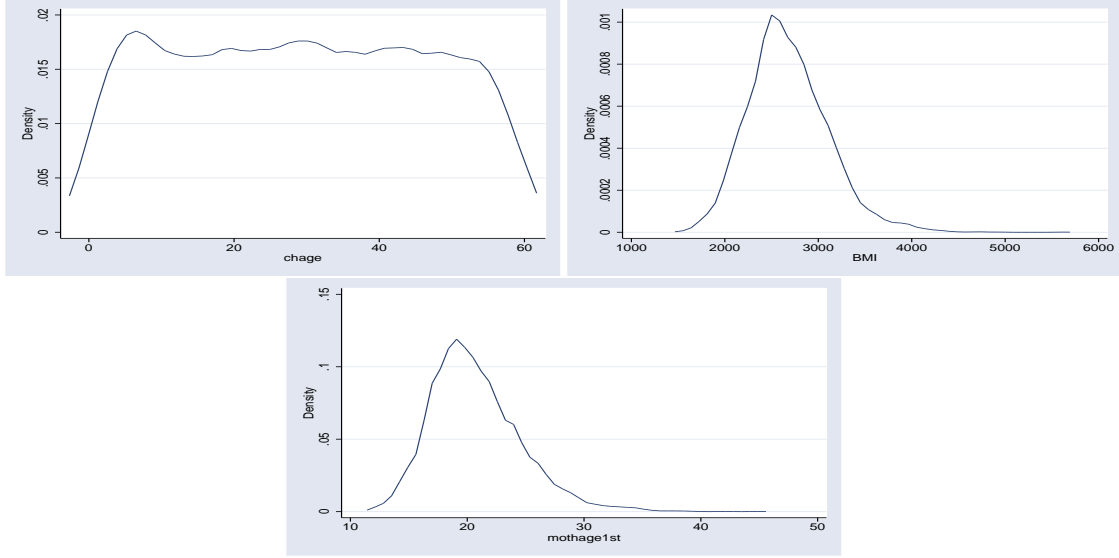


Figure 2: Density estimates of child's age (top left), kernel density estimates of Mother's body mass index (right left) and kernel density estimates of mother's age at birth in Egypt (bottom).

the household level. The range of overweight mothers is remarkably large, even within a region. For instance, 55 percent of mothers had overweight in Egypt. Figure 2 (right panel) shows that there are many overweight mothers (between 27-30 BMI) in the data set for Egypt.

Mother's age at birth (magb)

This is an important variable to fertility since it marks the onset of childbearing process. In a typical Middle Eastern culture, magb is expected to be highly correlated with age at first marriage. Delays in magb may indicate late establishment of marriage unions and hence implies shortening of the reproductive period and as consequence a reduced fertility. Figure 2 (Bottom panel) shows the kernel density estimates of mother's age at birth. It reflects the effect of the increasing age of mother at birth, a few mother fall in the age group (12-19). It indicates that most women had become mothers between age 18-24. Further, it shows that only a small percentage of women older than age 25 had given birth at the time of the survey.

Spatial covariates

Figure 1 shows that lower Egypt, essentially some districts in Nile Delta, are associated with significantly higher rate of illness. Red areas indicate that there is a negligible effect within these areas, green areas reflect a strong effect in

these regions and gray areas indicate that there are no children live in these regions according to the data set.

3 Bayesian Geoadditive Regression and Latent Variable Models

Geoadditive regression models extend (generalized) linear models for various types of response variables by adding nonparametric terms for nonlinear effects of continuous covariates and geographical 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.

3.1 Geoadditive Probit Regression

Let y_1, \dots, y_p denote p observable binary responses, such as the three disease indicators in our case study, and x_1, \dots, x_p corresponding covariate vectors. Note that some or even all components of these covariate vectors may be identical, thereby inducing association between the responses. Separate probit models with linear predictors can be defined through

$$P(y_j = 1|x_j) = \Phi(\beta_{0j} + x_j' \beta_j) \quad j = 1, \dots, p, \quad (1)$$

where Φ is the standard normal distribution function. Probit models can be based on Gaussian linear models

$$\tilde{y}_j = \beta_{0j} + x_j' \beta_j + \epsilon_j, \quad \epsilon_j \sim N(0, 1) \quad (2)$$

for unobservable auxiliary variables \tilde{y}_j through the threshold mechanism

$$y_j = 1 \Leftrightarrow \tilde{y}_j > 0, \quad y_j = 0 \Leftrightarrow \tilde{y}_j \leq 0. \quad (3)$$

Geoadditive probit models are obtained by extending the linear predictor

$\eta_j^{lin} = \beta_{0j} + x_j' \beta_j$ to a geoadditive predictor

$$\eta_j^{geo} = \beta_{0j} + x_j' \beta_j + f_1^j(z_1) + \dots + f_k^j(z_k) + f_{geo}^j(s).$$

The smooth functions f_1^j, \dots, f_k^j represent nonlinear effects of continuous covariates z_1, \dots, z_k . For simplicity, we only considered the case that these covariates are the same for each predictor $\eta_j^{geo}, j = 1, \dots, p$. The function f_{geo}^j represents the geographical effect of a spatial variable $s \in \{1, \dots, d\}$, indicating regions or districts in a country. The geographical effect $f_{geo}^j(s)$ of region s can be interpreted as a surrogate for unobserved variables with geographical information, incomplete or not covered by observable covariates. It may be split up into a structured part f_{str} for correlated spatial effects, and an unstructured part f_{unstr} for uncorrelated, local spatial effects, see section 3.3. Given the data $(y_{ij}, x_{ij}, z_{i1}, \dots, z_{ik}, s_i), i = 1, \dots, n$, where s_i is the region $\in \{1, \dots, d\}$ where individual i lives, geoadditive probit models for observations are given by

$$P(y_{ij} = 1 | \eta_{ij}^{geo}) = \Phi(\eta_{ij}^{geo}), \quad i = 1, \dots, n, \quad j = 1, \dots, p \quad (4)$$

$$\eta_{ij}^{geo} = \beta_{0j} + x'_{ij}\beta_j + f_1^j(z_{i1}) + \dots + f_k^j(z_{ik}) + f_{geo}^j(s_i).$$

Correspondingly, unobservable geoadditive Gaussian models for the auxiliary variables \tilde{y}_j are given by

$$\tilde{y}_{ij} = \eta_{ij}^{geo} + \epsilon_{ij}, \quad \epsilon_{ij} \quad i.i.d \sim N(0, 1). \quad (5)$$

The unknown parameters β_{0j}, β_j and functions $f_1^j, \dots, f_k^j, f_{geo}^j$ have to be estimated from the data. We follow a semiparametric Bayesian approach as developed in Fahrmeir and Lang (2001) and Brezger and Lang (2005). We assume diffuse, non-informative priors based on Markov Chain Monte Carlo (MCMC) techniques $p(\beta_{0j}) \propto const, p(\beta_j) \propto const$. Functions f_1, \dots, f_k follow P-spline priors, and the geographical effect f_{geo} is modelled through a Markov random field. Details about these priors are outlined in Khatab, 2007, and the MCMC inference is implemented in BayesX.

3.2 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, \dots, 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 latent variable models (LVM) is that the vector of the p observable variables can be represented, at least partly, by one or more latent factors or variables v with a lower dimension. As in our case study,

where we introduce the latent variable v “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 (2007) and Fahrmeir and Khatab (2007). The simplest LVM for Gaussian responses \widetilde{y}_j , $j = 1, \dots, p$, and scalar v is given through

$$\widetilde{y}_{ij} = \lambda_j v_i + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p, \quad (6)$$

with i.i.d, Gaussian errors ϵ_{ij} . In this model, v_i is the unobservable value of individuum i , λ_j is the “factor loading,” and $\lambda_j v_i$ is the effect of v_i . The restriction to $\sigma_v = \text{var}(v) = 1$ is necessary for identifiability reasons; otherwise λ_j would only be identifiable up to the constant $\sigma_v \neq 1$. If the \widetilde{y}_{ij} cannot be observed directly but only binary indicators

$$y_{ij} = 1 \Leftrightarrow \widetilde{y}_{ij} > 0,$$

then we obtain a probit LVM

$$P(y_{ij} = 1 | v_i) = \Phi(\lambda_j v_i) \quad i = 1, \dots, n, \quad j = 1, \dots, p. \quad (7)$$

One aspect of the latent variable is that it captures part of the variability of the responses. Secondly, although responses \widetilde{y}_{ij} or y_{ij} are conditionally independent for the given v_i , they are correlated marginally. These simple models can be extended to geoadditive probit LVMs as follows:

In the most general form, we augment the geoadditive predictors η_{ij}^{geo} in model (4) or (5) to

$$\eta_{ij}^{geo} + \lambda_j v_i, \quad i = 1, \dots, n, \quad j = 1, \dots, p, \quad (8)$$

resulting in the *measurement model*

$$\widetilde{y}_{ij} = \eta_{ij}^{geo} + \lambda_j v_i + \epsilon_{ij}, \quad (9)$$

with i.i.d errors $\epsilon_{ij} \sim N(0, 1)$ for the auxiliary variables \widetilde{y}_{ij} and in

$$P(y_{ij} | \eta_{ij}^{geo}, v_i) = \Phi(\eta_{ij}^{geo} + \lambda_j v_i), j = 1, \dots, p$$

for the binary responses.

Secondly, we allow that the latent variable v is influenced by covariates in form of a *geoaddivitive structural model*

$$v_i = u_i' \alpha + f_1(w_{i1}) + \dots + f(w_{iq}) + f_{geo}(s_i) + \delta_i, \quad (10)$$

with i.i.d. Gaussian errors $\delta_i \sim N(0, 1)$. For identifiability reasons as mentioned before it is assumed that $(\delta_i) = 1$, and that the predictor for v contains no intercept term. The additional covariates u, w_1, \dots, w_k and the location variable s act directly on the latent variable v , but indirectly on the observable responses. Covariates included in the structural model (10) must not be included in the measurement model at the same time, again for identifiability reasons. In particular, a spatial effect f_{geo} has to be included in either the measurement or the structural model. As for our application, we will restrict the attention to probit LVMs with linear predictors for the measurement model, i.e.,

$$P(y_{ij}|x_{ij}) = \Phi(\beta_{0j} + a_{ij}' \beta_j + \lambda_j v_i)$$

and geoaddivitive structural models (10) for v_i . The covariates a_j are different from the covariates u, w_1, \dots, w_k , and they have direct effects β_j on the observed responses. The effects α of u , and the nonparametric effects as well as the spatial effect are indirect effects. We used a_j (instead of x_j) as direct covariates in the case of latent variable model for simplicity.

3.3 Priors and Bayesian Inference

To complete the Bayesian model specifications, priors have to be assigned. For the direct effects β_{0j}, β_j and the indirect parametric effects α , we assume diffuse priors

$$p(\beta_{0j}) \propto \text{const}, \quad p(\beta_j) \propto \text{const}, \quad p(\alpha) \propto \text{const}.$$

For the factor loadings, we specify informative Gaussian priors

$$p(\lambda_j) \propto N(0, \sigma_j^2),$$

with $\sigma_j^2 = 1$ as the standard choice, to avoid the so-called Heywood cases (see e.g. Raach, 2005).

Priors for functions

For a function $f(w)$ of a continuous covariate w , we assume Bayesian P-spline-priors as in Brezger and Lang (2005).

Priors for spatial effects

We usually split $f_{geo}(s)$ into a smooth structured effect and an unstructured effect, i.e.

$$f_{geo}(s) = f_{str}(s) + f_{unstr}(s)$$

4 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 here, we used the Deviance Information Criterion (DIC) of Spiegelhalter et al.(2002) to select models in a formal way. Section 4.1 presents results of this first data analysis step. In the second step, we then applied geoadditive probit LVMs to analyze the data. While 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. It was decided to choose the covariates used in equation (9) for the measurement model, which have direct effects on the disease indicators; or in the case of the structural equation (10), those have indirect effects via their common impact on the latent variable "health status," we therefore proceeded more informally: 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. The results are presented in section 4.2.

4.1 Analyses with Separate Geoadditive Models

We present results for the following probit models, selected from a longer hierarchy of models. The responses y_j , $j = 1$ (diarrhea), 2 (fever), 3 (cough) are coded as

$$y_i = \begin{cases} 1 & \text{if child had disease 2 weeks prior to the survey} \\ 0 & \text{if not} \end{cases} \quad (11)$$

The following covariates were considered in the analysis in both countries:

Metrical covariates

Chage: Child's age in months.

BMI: Mother's body mass index.

Mageb: Mother's age at birth.

Categorical covariates (in effect coding)

male: Child's sex: male or female (reference category).

educ: Mother's educational attainment: incomplete primary, complete primary, and incomplete secondary school; or complete secondary school and higher education (reference category).

trepr: Whether mother had treatment during pregnancy: yes or no (reference category).

anvis: Whether mother had antenatal care: yes or no (reference category).

water: Source of drinking water: controlled water or no (reference category).

toilet: Has flush toilet at household: yes or no (reference category).

urban: Locality where respondent lives: urban or rural (reference category).

radio: Has a radio at household: yes or no (reference category).

elect: Has electricity: yes or no (reference category).

work: Mother's current working status: Working or not (reference category).

Spatial covariate

reg: Governorate where respondent resides.

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

M0: Included only district-specific effects.

$$M0 : \eta_{ij} = \beta_0 + f_{str}(reg) + f_{unstr}(reg) \quad (12)$$

M1: Includes all categorical covariates and the metrical covariates.

$$M1 : \eta_{ij} = \beta_{0j} + f_j(Chage) + f_j(BMI) + f_j(Mageb) + w'_i \gamma_j \quad (13)$$

M2: Adds district-specific effects to Model 1.

$$M2 : \eta_{ij} = \beta_{0j} + f_j(Chage) + f_j(BMI) + f_j(Mageb) + f_{str}(reg) + f_{unstr}(reg) + w'_i \gamma_j \quad (14)$$

$$M3 : \eta_{ij} = \beta_{0j} + f_j(Chage) + f_j(BMI) + f_j(Mageb) + f_{str}(reg) + f_{unstr}(reg) + z'_i \gamma_j \quad (15)$$

In these models, β_0 is a constant term and the covariate vector w in models M1 and M2 contains all the bio-demographic and health factors. In model M3 the

Model	Deviance	pD	DIC
Diarrhea			
M0	6364.46	15.45	6395.38
M1	5433.27	36.53	5506.34
M2	5432.74	36.91	5506.55
M3	5311.83	46.50	5404.84
Fever			
M0	7892.49	12.98	7918.47
M1	6972.51	36.74	7045.99
M2	6904.23	48.043	7000.32
M3	6911.25	44.38	7000.02
Cough			
M0	7076.87	14.43	7106.38
M1	6432.95	35.92	6504.78
M2	6330.83	48.96	6428.75
M3	6336.81	45.15	6427.11

Table 3: The Deviance Information Criterion (DIC)

vector w is reduced to the vector z by omitting factors of education, type of toilet and source of water. The metrical covariates child's age, mother's BMI and mother's age at birth are allowed to have a non-linear effect on the diseases of child as well as the spatial effects f_{str} and f_{unstr} . It turned out that model M3 for each type of diseases is superior in terms of the DIC.

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 shown in models M1, M2 and M3. The results for these models are given in Khatab (2007). However, we will only focus in this paper on the results of M3 which is the best model in terms of DIC. The results for the categorical covariates (M3) are shown in Tables 4 through 6 for the three diseases, respectively.

In figures 3-5 for the effects of the continuous covariates on the three diseases, respectively, and in figures 6-8, these suggest district variation in prevalence of diarrhea, fever, and cough using geospatial separate analyses.

Diarrhea

Table 4 displays the estimated categorical effects of these variables (male, urban, mother working status, mother had treatment during pregnancy, antenatal visit, availability of radio, availability of electricity, source of drinkable water, mother's education, and toilet facility) on diarrhea disease in Egypt. The results of Egypt indicate a significant impact of mother had treatment during pregnancy on disease of diarrhea and a significant impact of antenatal visit. However, antenatal visit has a positive effect. This analysis also suggests that mother's education, mother working status, toilet facility, availability of electricity and source of drinkable water have little or non significant effects.

With regard to the non-linear effects, figure 3 shows from top to bottom: the (nonlinear) effects of age of the child, mother's body mass index and mother's age at birth for model M3, respectively, modelled through Bayesian P-splines. The nonlinear effect of child's age suggested 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 the children of mothers whose have 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 larger than 43-45, there are few observations and the credible intervals gets wider. A somewhat higher risk for diarrhea seems to exist for mothers who have a BMI between 27 and 30, where a bump appears. In addition, we find the influence of mother's age (second panel from the bottom to the top bottom of figure 3) 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 age group (18-22 years) are at a higher risk of diarrhea compared to 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 (right panels) and its corresponding posterior mean

of unstructured random effects (left panels). For the model M3 for the diarrhea disease, the geographical pattern of regions in the right panel of figure 6 reflects the estimated posterior means of the structured random effects on diarrhea. Obviously, there exists a district-specific geographical 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,), 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 in the upper area in spite of being surrounded by some districts with lower risk.

Fever

The fixed parameters show that the prevalence of fever in Egypt (table 5) is higher among infants from mothers who are working, males, and children from mothers who obtained antenatal visits during pregnancy. Availability of radio in the household is associated with a lower risk of fever morbidity. On the other hand, the results suggest that mother's educational attainment, whether the mother received injection during pregnancy or not, availability of a flush toilet, availability of electricity, source of drinkable water and locality of residence have only a slight influence on fever morbidity in Egypt. In addition, mother who obtained antenatal care during pregnancy, had access to electricity and radio have a lower 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 in the three models in Egypt. 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. The effect of mother's BMI on fever is shown in figure 4. It is observed that mother's BMI has a slight significant impact on child health status in Egypt. Furthermore, it declines for mothers with a BMI of less than 20, and is less pronounced for mothers with BMI between 20-35, in spite of a blip between BMI of 30 and 35, which is caused by overweight mothers in Egypt, and over a BMI of 40, there are only few observations (wide credible interval). Unexpectedly, the effect of mother's BMI $f(BMI)$ in the three models turns out to be almost linear.

With regard to the non-linear effect of mother's age at birth on fever morbidity, the figure 4 displays that children from younger mothers (< 20 years) are at considerably higher risk of morbidity compared to children from mothers who are in the middle-aged group (25-35) and the impact of mother's age on fever disease is quite similar for both countries.

The geographical 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, and this variation could be

Variable	Mean	S.dv	2.5%	median	97.5%
const	-1.116*	0.169	-1.469	-1.119	-0.806
male	0.060*	0.019	0.021	0.060	0.097
urban	-0.062*	0.024	-0.109	-0.063	-0.016
work	0.010	0.025	-0.042	0.010	0.057
trepr	0.065*	0.031	0.002	0.065	0.129
anvis	0.080*	0.022	0.036	0.079	0.123
radio	-0.051	0.025	-0.101	-0.050	0.001
elect	-0.002	0.094	-0.177	-0.003	0.201

Table 4: Fixed effects of model(M3) on diarrhea-Egypt.

attributed to environmental risks, which in turn influence exposure to disease. The unstructured effects are similar to the structured effects. The gray area, however, indicates that no children live there.

Cough

The results indicate that children from mothers who attended an antenatal care during pregnancy, and currently working face a high rate of cough disease compared to children from mothers who are not working and did not attended any care. The results also suggested that ownership of radio facility has a negative impact on cough disease in Egypt. It is observed that the boys under 5 years are more likely to get cough morbidity than girls. The rest of categorical covariates have either a negligible impact or an insignificant effect on cough morbidity table (6). Further, the results indicate that some covariates such as availability of electricity, source of water, place of residence, and education attainment are only at the borderline to significance.

The non-linear effect of child's age for model M3 has a similar pattern to diarrhea and fever. The same is true for mother's BMI and mother's age at birth, for Egypt.

Spatial effect on cough in Egypt is seen in figure (8). The results suggest that significantly high rates of cough illness are associated with Damietta, Dakhalia and Esmaliya.

Variable	Mean	S.dv	2.5%	median	97.5%
const	-0.270*	0.186	-0.603	-0.275	0.095
male	0.046*	0.017	0.009	0.044	0.080
urban	0.006	0.021	-0.036	0.006	0.049
work	0.043	0.023	-0.003	0.045	0.089
trepr	0.025	0.030	-0.036	0.025	0.0868
anvis	0.075*	0.0198	0.039	0.074	0.115
radio	-0.069*	0.024	-0.119	-0.068	-0.019
elect	-0.211	0.167	-0.536	-0.200	0.104

Table 5: Fixed effects of model (M3) on fever-Egypt.

Variable	Mean	S.dv	2.5%	median	97.5%
const	-0.420*	0.199	-0.811	-0.420	-0.036
male	0.046*	0.017	0.010	0.046	0.077
urban	0.033	0.021	-0.009	0.032	0.072
work	0.0596*	0.0235	0.010	0.060	0.104
trepr	0.026	0.029	-0.027	0.0273	0.084
anvis	0.059*	0.020	0.019	0.059	0.099
radio	-0.066*	0.024	-0.115	-0.066	-0.017
elect	-0.090	0.177	-0.422	-0.093	0.270

Table 6: Fixed effects (M3) on cough-Egypt.

4.2 Discussion

Fixed Effects

As for child's gender, it is widely believed that probability of disease is higher for males due to biological reasons. Although, boys are noticeably more likely than girls to be taken to a provider for treatment (EDHS 2003). However, some studies show higher female mortality indicating gender discrimination. The results show that a child's gender is mostly significant and has a large impact on the three types of diseases in Egypt.

The effects of urban versus rural place of residence are different for the three diseases: For diarrhea, living in urban areas lowers the risk, for fever and cough the effect is not significant for children from urban vs.rural areas. These results support the important role of the public health policy in rural-urban disparities.

Mothers who attended a clinic to receive antenatal care during the period of pregnancy are expected to have lower problems in comparison to those who had not received any care. The results for Egypt, however, suggest the contrary: the factor antenatal visit has a positive effect on the indicators of disease! A possible reason could be that there are few of mothers who obtained antenatal visits frequently during their pregnancy. In addition, there are only 10% who had treatment during their pregnancy or maybe the reason of getting care was

not related to their pregnancy. The ownership of radio facilitates the acquisition of disease and vaccination information, allowing a more effective allocation of resources to produce child health. Therefore, it has a negative significant effect on the morbidity as suggested by the results for Egypt.

Concerning current working status of mother, these results suggest a significant effect of this variable on fever and cough morbidity in Egypt, however the effect is positive. The problem is when mothers engage in out-of-home employment it curtails the duration of full breastfeeding and necessitates recently introduced supplementary feeding, often by the illiterate care-takers, and that could have a side effect on the health of child in the early months.

Non-Linear Effects

In general, the results show that the risk of having diseases in the two-week reference period reaches its peak at 11 months 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 contributes to the lower prevalence of diarrhea. On the other hand, prevalence is quite high when the child has lost inborn immunity and when it 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 to mothers 20-35 years old. The results reflect a slight effect of mother's age at birth on the morbidity of children.

In the literature, the influence of the body mass index (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 at a low level equilibrium. The higher impact of BMI through the interval between 27-30, 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

Parameter	Mean	Std	2.5%	97.5%
Factor Loadings				
1. Fever λ_{11}	2.2	0.34	1.78	3.03
2. Cough λ_{21}	0.87	0.04	0.77	0.959
3. Diarrhea λ_{31}	0.67	0.03	0.616	0.73

Table 7: Results of Model LVM0 for Egypt with $\eta = 0$.

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 higher influence on child in the north-east part, affecting the most of districts there. Food insecurity associated with water supplies and quality of water could be a reason for these negative effects in this area.

5 Analyses with Latent Variable Models

As previously discussed in section 3, we now investigate how the three diseases can be interpreted as indicators of a latent variable v "health status" of children, how much of the variation of v can be explained through a geoaddivitive 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 a correlation among disease indicators. To demonstrate the latter property, we first consider a classic model without any covariates, i.e. in turns of auxiliary variables.

(LVM0):

$$P(y_{ij} = 1|v_i) = \Phi(\lambda_j v_i), \quad v_i \sim N(0, 1) \quad (16)$$

and $\eta = 0$, so that $v_i \sim N(0, 1)$. Table 7 shows the estimates for the factor loadings λ_j , $j = 1, 2, 3$ implying considerable (marginal) correlation.

Our next model is selected on the basis of the separate analyses as explained at the beginning of this section. This leads to the latent variable model

$$P(y_{ij}|x_{ij}) = \Phi(\beta_{0j} + a'_{ij}\beta_j + \lambda_j v_i), \quad j = 1, 2, 3$$

with the structural model

$$v_i = u'\alpha + f_1(Chage_i) + f_2(BMI_i) + f_3(Mageb_i) + f_{geo}(reg_i) + \delta_i$$

for the latent variable. The vector a' (measurement model) comprises the covariates with direct effects (such as urban, availability of electricity and controlled water in LM1 for Egypt) on y_j , and u comprises the remaining categorical covariates (such as sex, mother's education, etc. in LM1 for Egypt) having common effects on the latent variable v . Because the patterns for the nonparametric functions and the spatial effects were rather similar in the separate analyses, they were included in the geosadditive predictor for v .

The results of latent variable models for categorical covariates are in table 8. Factor loadings are slightly lower than for the factor analysis without covariates.

Because indirect effects affect the latent variable, they cover a larger range of values and thus exert more influence on the variability of the indicators, even if the factor loadings are slightly lower.

The results show that the parametric indirect covariates of male, antenatal visit, having radio, and mother's working status have a significant effect on the latent variables. The results indicate that the mother's education, ever had treatment during pregnancy and toilet facility have only a non-significant or slight effect on the latent variables. Concerning the categorical direct covariates, the results indicate a significant effect of urban on cough and diarrhea. However, the effect of urban on cough is positive. The reason is that only one instead of three separate effects have to be estimated. The results of LVM1 are quite consistent with the previous results which are obtained using geosadditive models for each kind of disease in section 4.1. The insignificant (access to electricity and water) parametric direct covariates were included in the parametric indirect effects in LVM2 (it is not included in this paper) and they still have nonsignificant impacts on the indicators of health status in Egypt, therefore, we excluded these covariates in model LVM3 (table 9).

The pattern for non-linear 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. Therefore, only the results of model LVM2 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 under 5 living in these areas.

5.1 Conclusions

In this paper we investigated socioeconomic and public health, and spatial determinants of morbidity, measured through prevalence of fever, cough and diarrhea, in Egypt. Our analyses show that geoaddivitive models are needed to adequately assess nonlinear covariate effects and geographical effects within a joint model. With traditional regression model these effects are difficult if not impossible to model and to detect. 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.

We conclude by pointing out some conceptual and technical problems associated with information on prevalence of fever, 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 have died before the survey, and whether the cause of dying was kind of the diseases which 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 socio-economic groups because perception of illness is not the same across different social groups. Third, loss of memory of events as well as misinterpretation of the reference period can also contribute to the problems associated with the prevalence of diarrhea (Bateman and Smith, 1991; Gaminiratne, 1991).

Acknowledgment

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

References

- BayesX (2005). Software for Bayesian inference by Brezger, A., Kneib, T., Lang, S.
- Besag, J., York, Y. and Mollié, A. (1991). Bayesian Image Restoration with Two Applications in Spatial Statistics (with Discussion). *Ann. Inst. Statist. Math.*, 43, 1-59.
- Brezger, A. and Lang, S. (2005). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis* 50, 967-991.
- Demographic and Health Survey (EDHS) for Egypt (2003).
- Heywood, H.B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society A* 134, 486-501.
- Fahrmeir, L., and Khatib, K. (2007). Geoadditive Latent Variable Modelling of Child Morbidity and Malnutrition in Nigeria, Discussion paper 20, Department of Statistics, Ludwig-Maximilians-Universität München.
- Fahrmeir, L. and Lang, S. (2001). Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Applied Statistics (JRSS C)*, 50, 201-220.
- Fahrmeir, L., and Raach, A. (2007). A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*, 72, 327-346.
- Kandala, N.B. (2002). Spatial Modelling of Socio-economic and Demographic Determinants of Childhood Undernutrition and Mortality in Africa. Ph.D Thesis, Shaker Verlag, Munich, Germany.
- Kandala, N.B., Lang, S., Klasen, S. and Fahrmeir, L. 2001. Semiparametric Analysis of the Socio-Demographic Determinants of Undernutrition in Two African Countries. *Research in Official Statistics*, Vol. 4 No.1:81-100.217.
- Kandala, N.B., Chen Ji, Nigel Stallard, Saverio Stranges, and Francesco P. Cappuccino. Spatial Analysis of Risk Factors for Childhood Morbidity in Nigeria. *Journal of Tropical Medicine*, 77(4), 2007, pp.770-778.
- Khatib, K. (2007). Analysis of childhood diseases and malnutrition in developing countries of Africa. Ph.D thesis, Dr. Hut Verlag, Munich, Germany.
- Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13, 183-212.
- Lang, S.; Brezger, A.; Kneib, T. (2005). BayesX: Analysing Bayesian structured additive regression models. *Journal of Statistical Software*, 14 (2005), Heft 11.

- Langsten and Hill (1994). Diarrhoeal Disease, Oral Rehydration, and Childhood Mortality in Rural Egypt. *Journal of Tropical Pediatrics* 1994 40(5):272-278; doi:10.1093/tropej/40.5.272.
- Miller, P. and Hirschhorn, N. (1995). The Effect of Nutritional Control of Diarrhoea Diseases Program on Mortality: The Case of Egypt. *Social Science Medicine*. Vol.40, No.10, pp.s1-s30.
- Miller P, Loza S, Terreri N, Nagaty A, Bayoumi M, Hirschhorn N, Gipson R. United Nations Children's Fund, Cairo, Egypt. Diarrhoea and mortality in Menoufia, Egypt. *J Diarrhoeal Dis Res* 1994 Dec;12(4): 293.
- Moustaki, I., Jeoreskog, K.G. and Mavridis, D.(2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables:a comparison of LISREL and IRT approaches. *Structural Equation Modeling* 11, 487-513.
- Spiegelhalter D., Best N., Carlin B., and Van der Line A. (2002). Bayesian Measures of Models Complexity. *Journal of the Royal Stat. Soc. B*(64):1-34.
- Raach, A.W. (2005). A Bayesian semiparametric latent variable model for binary, ordinal and continuous response. Dissertation, available from edoc.ub.uni-muenchen.de

Parameter	Mean	Std	2.5%	97.5%
Factor Loadings				
1. Fever λ_{11}	1.29*	0.093	1.12	1.46
2. Cough λ_{21}	0.82*	0.04	0.75	0.91
3. Diarrhea λ_{31}	0.79*	0.04	0.71	0.87
Parametric Indirect Effects				
male	0.135*	0.038	0.059	0.208
anvis	0.218*	0.044	0.131	0.30
trepr	0.088	0.061	-0.034	0.208
work	0.123*	0.05	0.023	0.22
radio	-0.169*	0.05	-0.269	-0.070
educ	-0.061	0.032	-0.125	0.001
toilet	-0.129	0.094	-0.319	0.051
Semi-Parametric Indirect Effects				
Chage	0.059*	0.043	0.014	0.169
BMI	0.017*	0.028	0.000	0.085
Mageb	0.004*	0.011	0.0003	0.019
reg	0.201*	0.112	0.063	0.484
Parametric Direct Effects				
urban(a_{11})	0.0329	0.07	-0.1	0.17
elect(a_{12})	-0.36	0.274	-0.89	0.182
water(a_{13})	0.129	0.089	-0.041	0.30
urban(a_{21})	0.152*	0.054	0.044	0.25
elect(a_{22})	-0.071	0.23	-0.52	0.39
water(a_{23})	-0.016	0.071	-0.16	0.124
urban(a_{31})	-0.207*	0.056	-0.317	-0.09
elect(a_{32})	0.012	0.22	-0.42	0.47
water(a_{33})	0.042	0.072	-0.098	0.187

Table 8: Results of LVM1 including direct and indirect effects for Egypt.
(*: Statistically significant at 2.5%)

Parameter	Mean	Std	2.5%	97.5%
Factor Loadings				
1. Fever λ_{11}	1.273*	0.099	1.090	1.487
2. Cough λ_{21}	0.824*	0.040	0.746	0.911
3. Diarrhea λ_{31}	0.796*	0.047	0.706	0.889
Parametric Indirect Effects				
male	0.135*	0.038	0.060	0.209
anvis	0.22*	0.044	0.138	0.313
work	0.127*	0.050	0.02	0.225
radio	-0.186*	0.049	-0.286	-0.090
educ	-0.065	0.033	-0.129	0.008
Semi-Parametric Indirect Effects				
Chage	0.0597*	0.043	0.014	0.175
BMI	0.016*	0.027	0.0008	0.088
Mageb	0.003*	0.0056	0.0003	0.001
reg	0.202*	0.106	0.069	0.0473
Parametric Direct Effects				
urban(a_{11})	0.041	0.066	-0.088	0.171
urban(a_{21})	0.141*	0.055	0.0318	0.248
urban(a_{31})	-0.209*	0.056	-0.316	-0.09

Table 9: Results of LVM3 including direct and indirect effects for Egypt

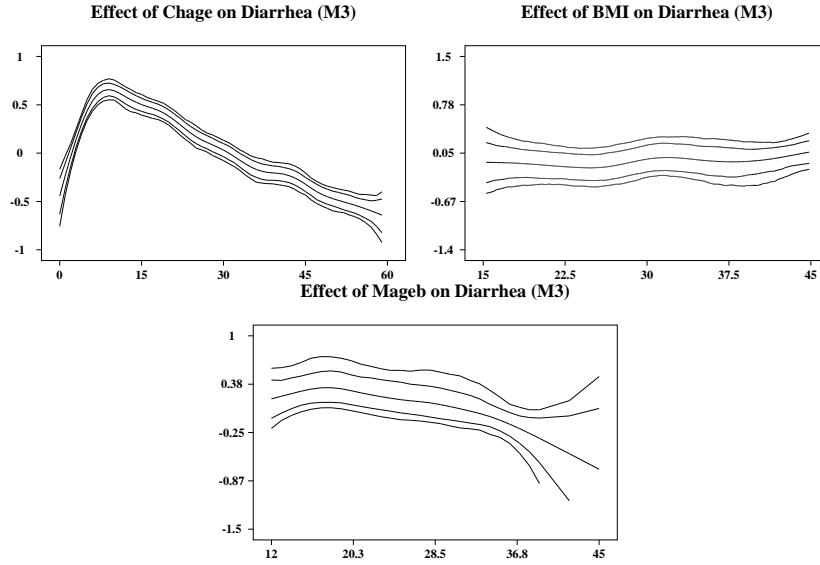


Figure 3: Non-linear effects of child's age, mother's BMI and mother's age for M3 on diarrhea for Egypt using probit model.

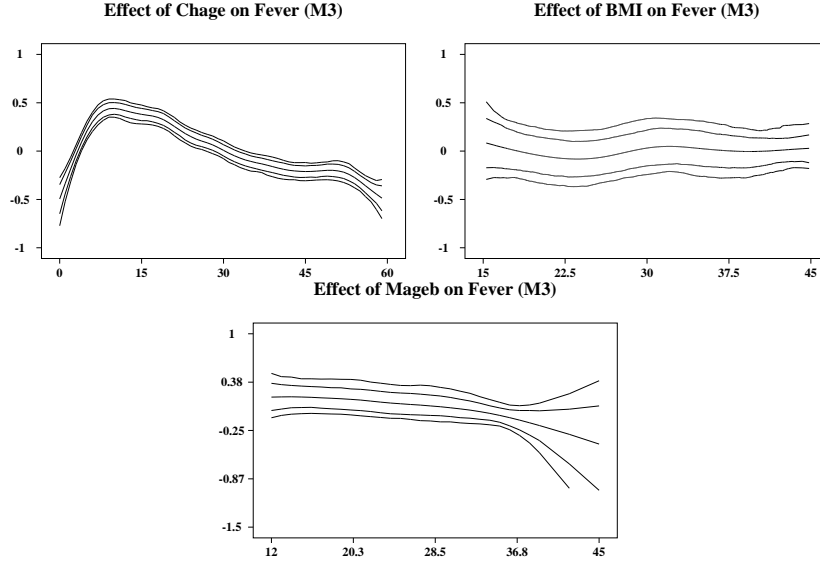


Figure 4: Non-linear effects of child's age, mother's BMI, and mother's age (for M3) on fever for Egypt using probit model.

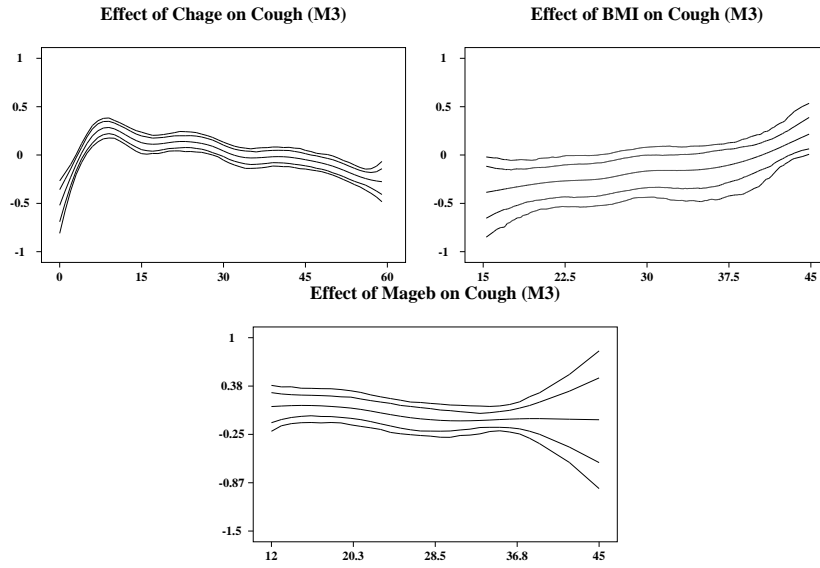


Figure 5: Non-linear effects of child's age, mother's BMI, and mother's age (for M3) on cough for Egypt using probit model.

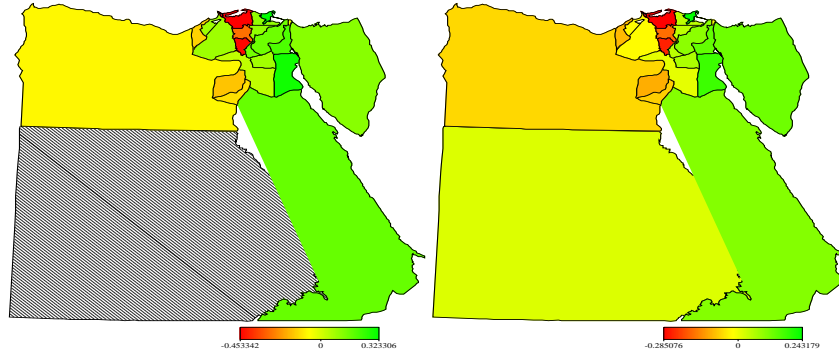


Figure 6: Maps of Egypt for diarrhea showing unstructured (left panel) and structured (right panel) spatial effects for M3 using probit model .

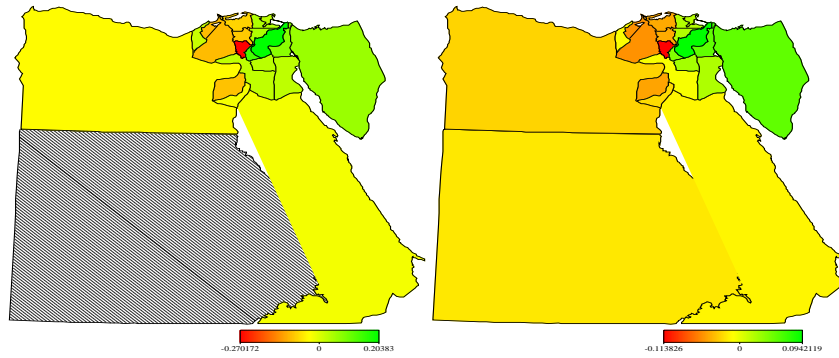


Figure 7: Maps of Egypt for fever showing unstructured (left panel) and structured (right panel) spatial effects for M3 using probit model.

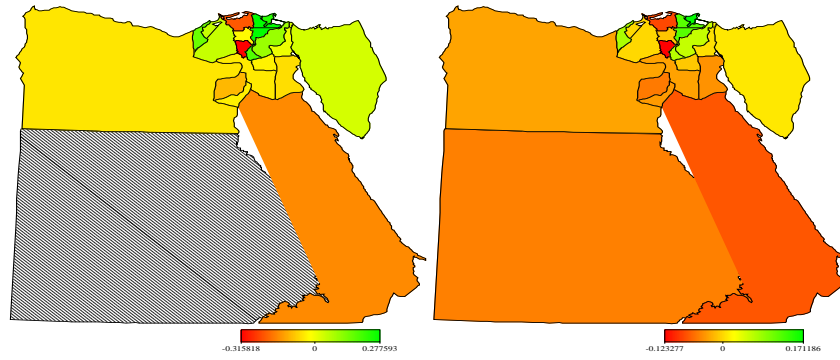


Figure 8: Maps of Egypt for cough showing unstructured (left panel) and structured (right panel) spatial effects for M3 using probit model .

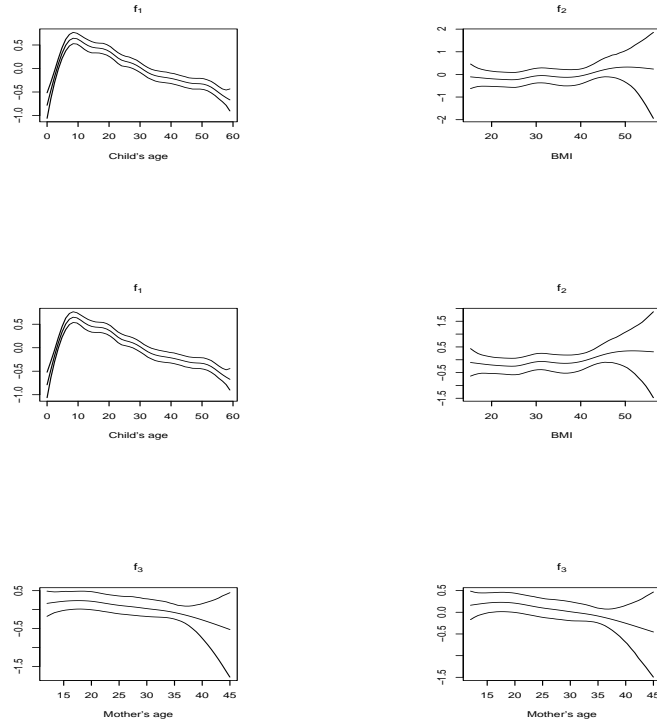


Figure 9: Non-linear effects from top to bottom: child's age , mother's BMI and mother's age at birth (for model LVM1), child's age, mother's BMI and mother's age at birth (for model LVM3) on the indicators of a latent variable "health status" of children disease for Egypt using Bayesian latent variable model for binary responses.

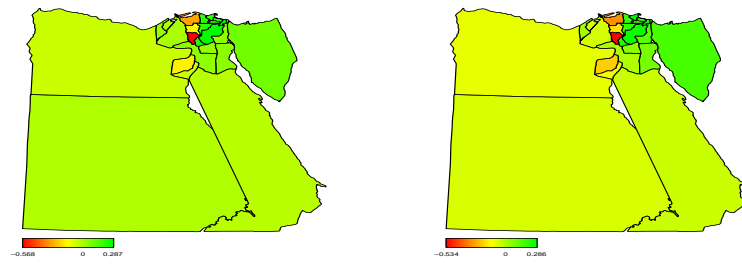


Figure 10: Posterior mean for latent variable model for LVM1 (left panel) and LVM3 (right panel) on diseases in Egypt.