LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Samson B. Adebayo, Ludwig Fahrmeir & Christian Seiler

Geoadditive Latent Variable Modelling of Count Data on Multiple Sexual Partnering in Nigeria

Technical Report Number 066, 2009
Department of Statistics
University of Munich

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

Geoadditive latent variable modelling of count data on

multiple sexual partnering in Nigeria

Samson B. Adebayo[*], Research & Evaluation Division, Society for Family Health, Abuja, Nigeria

Ludwig Fahrmeir, Department of Statistics, Ludwig-Maximilians-University of Munich, Munich, Germany

and

Christian Seiler, IFO Institute for Economic Research at the University of Munich, Munich, Germany.

* Corresponding author: sadebayo@sfhnigeria.org, adebayo_sba@yahoo.com

Summary

The 2005 National HIV/AIDS and Reproductive Health Survey in Nigeria provides evidence that multiple sexual partnering increases the risk of contracting HIV and other sexually transmitted diseases. Therefore, partner reduction is one of the prevention strategies to accomplish the Millenium development goal of halting and reversing the spread of HIV/AIDS. In order to explore possible association between sexual partnering and some risk factors, this paper utilizes a novel Bayesian geoadditive latent variable model for count outcomes. This allows us to simultaneously analyze linear and nonlinear effects of covariates as well as spatial variations of one or more latent variables, such as attitude towards multiple partnering, which in turn directly influences the multivariate observable outcomes or indicators. Influence of demographic factors such as age, gender, locality, state of residence, educational attainment, etc., and knowledge about HIV/AIDS on attitude towards multiple partnering is also investigated. Results can provide insights to policy makers with the aim of reducing the spread of HIV and AIDS among the Nigerian populace through partner reduction.

Key words: factor loading; geographical variations; latent variable model; MCMC; Nigeria; semiparametric Poisson model;

1. Introduction

The HIV and AIDS sentinel survey among pregnant women attending antenatal clinics in Nigeria revealed that the HIV prevalence rate has reduced from 5.8% in 2001 (the peak) to 4.4% in 2005 (Federal Ministry of Health (FMOH) [Nigeria], 2006a). However, substantial state level variation exists as prevalence ranges from 1.6% to 10.0% throughout the country. Nigeria still has a generalised epidemic with infection rates in high-risk groups being as high as 31.0% (FMOH, 2008a). Various programmes to mitigate the impact of HIV infection were put in place by the Federal Government of Nigeria in collaboration with partners to accomplish the Millenium development goals (MDGs) on halting and reversing the spread of HIV/AIDS by 2015.

Heterosexual intercourse with multiple partners, such as spouse, boy- or girlfriend, casual and commercial partners, has significant implication for sexual and reproductive health, including transmission of HIV and other sexually transmitted infections (STIs). The 2005 National HIV/AIDS and Reproductive Health Survey (NARHS) in Nigeria provides empirical evidence of this fact (see FMOH [Nigeria], 2006b). Therefore, partner reduction is one of the prevention strategies aimed at reducing the spread of HIV and AIDS.

In sub-Saharan African, and many other developing countries, polygyny inhibits the impact of HIV prevention and facilitates the spread of HIV. For example, see the study of Morris (2002) for Uganda. Following the recommendation of the World Health Organisation (WHO), condoms should be used consistently during casual and commercial sexual intercourse (Adetunji and Meekers, 2001). However, condoms are seldom used consistently in longer-term relationships in which there is a sense of commitment and trust (Flood, 2003;

Meekers, Klein, and Foyet, 2003; and Hearst and Chen, 2004). Therefore, in addition to consistent condom use and other HIV prevention approaches in Africa, concerted public health efforts should be directed at addressing the dangers of having multiple sexual partners at a time. These might have important implications for HIV prevention.

In this paper, we use data from the 2005 NARHS in Nigeria to investigate the influence of personal and demographic factors such as age, age at first sex, place of residence, knowledge about modes of transmission and prevention of HIV/AIDS, etc. on attitude towards multiple partnering among heterosexual men. We consider attitude towards multiple partnering as a latent variable, with numbers of different types of sexual partners as observable indicators. For analyzing NARHS data on these factors and indicators, we use a semiparametric latent variable model (LVM) for count indicators, see the technical report of Fahrmeir and Steinert (2006). Conceptually, this count data LVM is based on geoadditive LVMs for continuous and categorical indicators (Raach 2005; Fahrmeir and Raach, 2007), allowing to simultaneously analyze linear, non-linear and geographical effects of covariates on latent variables.

The article is organized as follows: Section 2 describes the data in more detail. Section 3 outlines the statistical methodology. Data analyses and discussion of results are presented in Section 4. The paper is concluded in Section 5. An appendix provides more technical details on MCMC inference for the count data models of Section 3.

2. The NARHS Data

NARHS was the first nationally representative survey on HIV and AIDS in Nigeria. Selection of the eligible respondents (male aged 15-64 years and female aged 15 to 49 years) was based on a probability multi-stage sampling technique. The survey protocol was developed and managed by the technical committee (TC) and survey management committee (SMC) while ethical approval was granted by the appropriate institutional review board. Both written (for the literate respondents) and verbal consents with thumb printing (non-literate respondents) were sought from the eligible respondents. As contained in the survey protocol, confidentiality of information provided by the respondents was emphasised and ensured. For detailed description of the survey protocol, see FMOH [Nigeria] (2006b). For the purpose of this paper, a database for heterosexual males aged 15 to 64 years, and who have had sex in the last twelve months prior to the survey, was created from the main data.

In this paper, number and types of different sexual partners constitute the observable outcome variables considered as indicators for attitude towards multiple partnering. Information about types (spouse/cohabiting, boy/girl friend, casual and commercial) and number of sexual partners a respondent had in the last 12 months was obtained from the sexual history section of the NARHS questionnaire. The questionnaire also contains information about sexual behaviours, questions on knowledge about HIV/AIDS, as well as personal and demographic characteristics.

We explore influence of the following personal and demographic variables on multiple sexual partnering: respondent's age (in years) as at the time of the survey, reported age (in years) at first sex, length of stay in the place of domicile, marital status, educational attainment, religion, locality (rural/urban), state in Nigeria where the respondent lives, and whether respondent had been away from home for more than a month in the last twelve months preceding the survey. Further covariates included in the NARHS study were knowledge about mode of transmission and mode of prevention of HIV/AIDS, knowledge about sexually transmitted infections (STI), and knowledge that AIDS has no cure. Table 1 presents detailed description of the variables included in the analysis. Of the 4,962 male respondents that participated in the survey, only 3,174 have had sex in the last 12 months prior to the survey. The mean age of respondents was estimated at about 35.2 years with a standard deviation of 11.7 years. Information on age at first sex was available for only 2320 respondents with a mean age at first sex of 19.7 years with a standard deviation of about 4.2 years. Information on length of stay in a particular area where the respondent was living as at the time of the survey was available for 3,083 respondents, with a mean of 23.7 years and a standard deviation of 14.5 years. Numbers of different sexual partners vary considerably with spousal and cohabiting partners ranging from 0 to 10, number of girl friends ranging from 0 to 7, number of casual partners ranging from 0 to 8 and number of commercial partners ranging from 0 to 6. However, about 92.9% of the respondents had sex with spousal/cohabiting partners, 71.1% had sex with girl friends, 62.1% had sex with casual partners and 62.3% had sex with commercial partners. Due to missing observations in some of the covariates, 1820 observations were included in the final analysis. Figure 1 displays states in Nigeria with their locations.

3. Geoadditive latent variable models for count indicators

Our analysis is based on a flexible geoadditive latent variable model (geoLVM), where observed outcomes or indicators are count variables y_j, j=1,…,p. In the NARHS study, we consider up to p=4 indicators: number of spousal/cohabitating partners, number of boy/girl friends, number of casual partners, and number of commercial partners. Generally, LVMs consist of a measurement model for the vector $y=y_1, \ldots, y_p$ of indicators, conditional on one or more common latent variables or factors $v_1,...,v_q$, where $q<p$, and a structural regression model relating latent variables to a vector of covariates. Our geoLVM follows the lines of Fahrmeir and Raach (2007), where y is a vector of mixed Gaussian and categorical indicators and the structural model is a semiparametric geoadditive regression model.

3.1 *Measurement model*

Let y_{ij} denote the observed value of the indicator y_j, j=1,…, p, and v_i the unobservable value of a scalar (q=1) latent variable v, for individual i=1,…,n. Conditional on v_i, the basic measurement model used in our analysis is a log-linear Poisson model

$$y_{ij} \mid \mu_{ij} \sim Po(\mu_{ij}), \;\; \mu_{ij} = \exp(\alpha_j + \lambda_j v_i), \tag{1}$$

i=1, …, n, j=1,…, p. In (1), α_j is an intercept term, and λ_j is a 'factor loading' indicating the strength of relationship between the latent variable and the count indicator y_j. In our study, we think of the latent variable as 'attitude towards multiple partnering'. Model (1) can be extended by incorporating covariates and more than one (q>1) factors in the predictor. Then, the log-linear model for the rate μ_{ij} is generalized to

$$\mu_{ij} = \exp(\alpha_j^{'} u_i + \lambda_j^{'} v_i), \tag{2}$$

where u_i is a vector of covariates with effects α_j (including an intercept α_{j0}) $v_i = (v_{i1},...,v_{iq})'$, $q<p$, is a vector of (values of) latent variables, and $\lambda_j = (\lambda_{j1},...,\lambda_{jq})'$ is a corresponding vector of factor loadings. Note that (1) and (2) extend the usual linear predictor of log-linear models by adding the linear effects $\lambda_j^{'} v_i$ of common latent variables.

3.2 *Structural model*

Usually, structural models relate latent variables to a covariate vector x_i through a linear model. For one latent variable, this is a linear Gaussian regression

$$v_i = x_i^{'}\beta + \delta_i, \quad i=1, \ldots, n, \tag{3}$$

with *i.i.d.* errors $\delta_i \sim N(0,1)$. For identifiability reasons, the linear predictor must not contain an intercept term and the error variance is set to 1. The latent linear model (3) assumes that effects of covariates can be represented in a linear form. However, in our study continuous covariates, such as age, age at first sex, and length of stay in a locality are supposed to have nonlinear effects. Moreover, we want to explore geographical effects by including the state of Nigeria, where the respondent lives, as a spatial covariate. Therefore, we extend the linear regression model (3) to a geoadditive regression model

$$v_i = x_i^{'}\beta + f_1(z_{i1}) + ... + f_r(z_{ik}) + f_{geo}(s_i) + \delta_i \qquad (4)$$

where $f_1,...,f_k$ are nonlinear functions for the effects of additional continuous covariates $z_1,...,z_k$ and $f_{geo}(s_i)$ is the geographical effect of area or state $s_i \in \{1,...,S\}$, indexing S geographical regions such as the S=36 states of Nigeria plus the Federal Capital Territory. To assure identifiability, functions are centred about zero. The geoadditive model (4) has the same form as in the semiparametric LVM with measurement models for mixed Gaussian and categorical responses in Fahrmeir and Raach (2007), and it can be extended to *q*>1 latent variables, in complete analogy.

Together, the measurement model (1) or (2) and the structural model (3) define our geoLVM for count indicators. Inserting the geoadditive structural model (4) into the measurement model (1), we see that the covariates *x* and *z* and the spatial covariate *s* have common but only indirect impact on the observable indicators $y_j, j = 1,...,p$, through the common latent variable l, with factor loadings λ_j acting as weights. The common latent variable also automatically induces correlation between the indicators.

3.3 *Priors*

For Bayesian inference, which is the most natural conceptual approach for LVMs, we have to specify priors for unknown parameters and functions in (1), (2) and (4). For simplicity we focus on the special case of only one (q=1) factor. We proceed as in Fahrmeir and Raach (2007) and assume noninformative flat priors $p(\alpha_j) \propto const, p(\beta) \propto const$ for the intercept terms in (1), direct effects in (2) and the regression coefficients in (4).

For the factor loadings λ_j we choose informative (truncated) Gaussian priors to prevent so-called 'Heywood cases'. A Heywood case appears when a factor loads up completely on one indicator, which is highly implausible. The standard normal prior for factor loadings is a recommended standard choice in a Bayesian setting (see e.g. Lopes and West, 2004; Quinn, 2004). To ensure identifiability, we assume $\lambda_1 > 0$, i.e. the prior is truncated normal $\lambda_1 \sim N(0,1)I(\lambda_1>0)$, and $\lambda_j \sim N(0,1)$, $j = 2,\ldots,p$. The extension to more than one factor is described in Fahrmeir and Raach (2007).

Priors for functions $f_1,...,f_k$ of continuous covariates are defined through Bayesian P-splines, based on Lang and Brezger (2004) and Brezger and Lang (2006). Omitting indices, each function f is represented or approximated through a linear combination

$$f(z) = \sum_{l=1}^{L} \gamma_l B_l(z)$$

of B-spline basis functions. Smoothness of function f is achieved by penalizing differences of coefficients of adjacent B-splines (Eilers and Marx, 1996) or, in our Bayesian approach, by assuming first or second order Gaussian random walk smoothness priors

$$\gamma_l = \gamma_{l-1} + u_l \quad \text{or} \quad \gamma_l = 2\gamma_{l-1} - \gamma_{l-2} + u_l,$$

with *i.i.d.* errors $u_l \sim N(0,\tau^2)$. The variance τ^2 controls the smoothness of f. Assigning a weakly informative inverse Gamma prior $\tau^2 \sim IG(\varepsilon,\varepsilon)$, ε small, it is estimated jointly with the basis function coefficients.

For the geographical effects $f_{geo}(s), s = 1,...,S$, we assume a Gaussian Markov random field prior. Basically, this is an extension of first order random walk priors to two-dimensional spatial arrays, see Rue and Held (2005) for general information and Fahrmeir and Raach (2007) in the context of geoLVMs.

Full Bayesian inference is carried out via Gibbs sampling in combination with an auxiliary Gaussian mixture variable approach for Poisson responses, suggested in the context of count time series in Frühwirth-Schnatter and Wagner (2006). Details of all Gibbs steps are described in Fahrmeir and Steinert (2006); an outline is given in the appendix.

3.4 *Model choice issues*

For LVMs with continuous and categorical indicators and a linear structural model, Sammel, Ryan, and Legler (1997) provide motivation and some guidance on which covariates might be kept in the measurement model, and on which covariates should be relegated to the structural model. For classical Gaussian factor analysis, Lopes and West (2004) empirically study Bayesian model assessment based on Bayes factors, DIC and reversible jump MCMC. Currently, there are no automated purely data driven tools for model checking and diagnostics available for deciding on this and on other model choice issues in the semiparametric LVMs for count indicators considered here. A (conditional) version of the deviance information criterion (DIC) can be computed from the MCMC output, but its properties for model choice in LVMs are not well studied yet. More generally, development of formal model assessment in complex hierarchical models, in particular in LVMs, is desirable but just at the beginning.

Thus, model choice issues are based on substantive reasoning in combination with more informal statistical arguments. For example, 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 through common effects acting via the latent variables.

Our current strategy is as follows: We fit separate (univariate) geoadditive regression models to each of the outcomes y_j, j=1,…, p, using mainly DIC for choosing between competing models such as inclusion of covariates or deciding between linear and nonlinear effects. Then we relegate covariates with similar effects or patterns to the structural model while the rest is kept in the measurement model.

4. Data analysis and results

4.1 *Analysis*

To explore the impact of personal, demographic and risk factors on the attitude towards multiple sexual partnering, we based our analyses on all covariates described in Table 1. Therefore, only the following issues were of relevance for the model building process: Are there covariates which should be kept in the measurement model? Is it reasonable and possible to explain the impact of remaining covariates on all four types of partnering through one latent factor only? Because the types spouse/cohabiting partner, and perhaps also girl friend, seem to be different from casual and commercial partners with regard to HIV/AIDS risk factors, the following question arises: Should we base our analysis on two latent factors instead of one? Or should we consider spouse/cohabiting partners separately while applying a geoLVM to the remaining types only?

To deal with these issues, we followed the strategy for model choice as outlined in Section 3.4. Separate analyses for each of the four types of sexual partners revealed that marital status has great and rather different impact on the four types and should rather be considered as an 'offset' variable to be included in the measurement model. It also turned out that the effect of the continuous covariate length of stay could be assumed as linear. Furthermore, the impact of AIDS/HIV related factors was much less pronounced or even not significant for the number of spouse/cohabiting partners, and the pattern of nonlinear effects was somewhat different from corresponding patterns for other types. This observation was confirmed by a first attempt to analyze the data with all four indicators and one common latent factor: The factor loading on spouse/cohabiting factor was close to zero, indicating that the factor had only significant influence on the numbers of sexual partners for the remaining three types. On the other hand, analysis with two latent factors led to obvious identification problems. As a conclusion, we decided to consider only the numbers of the types y_1 girl friend, y_2 casual partner, and y_3 commercial partner as indicators for one common latent factor 'attitude towards multiple sexual partnering'. This results in a geoLVM with the measurement model

$$y_j \mid \mu_j \sim P_0(\mu_j),\ \mu_j = \exp(\alpha_{j0} + \alpha_{j1}\mathrm{fm} + \alpha_{j2}\mathrm{cm} + \lambda_1 l),\ \ j = 1,2,3,$$

where α_{j0} corresponds to the effect of the reference category 'never married' and α_{j1}, α_{j2} are the (additional) effects of the 0,1 dummy variables *fm*, *cm* for the categories 'formerly married' and 'currently married'. The predictor for the final structural model is

$$\eta = f_1(\mathrm{age}) + f_2(\mathrm{agefirstsex}) + f_{geo}(\mathrm{state}) + \beta_1 \cdot \mathrm{awayfromhome} + \ldots + \beta_{13} \cdot \mathrm{lengthstay}\,.$$

Appropriate priors as discussed in Section 3.3 are assumed on all unknown parameters and functions. For instance, for all nonlinear effects, cubic Bayesian P-splines with 5 knots was assumed. To estimate the smoothing parameters for non-linear and spatial effects, highly dispersed but proper inverse gamma hyper-priors are assigned to them. Hyperparameters a, b, were varied systematically. Results were found to be similar. Therefore, for this case-study, inverse gamma priors for the variance components with hyperparameters *a*=*b*=0.001 were used.

4.2 *Results*

Tables 2, 3 and 4 present findings for the factor loadings, direct and indirect parametric effects respectively. Shown are the posterior means, standard deviations and the 95% credible intervals for each parameter. From Table 2, we see that the latent factor loads up significantly on all three indicators. However, factor loadings are much higher for commercial and casual partners. This has to be taken into account when interpreting parametric and nonparametric effects of the structural model for y_1, y_2, y_3. The indirect linear effects of the structural model for the latent variable are shown in Table 4. To interpret the effects of the model for y_1, multiply the factor loadings for y_1, y_2 and y_3 respectively, by the value for a particular fixed effect. For instance, the effect of *knowledge that AIDS has no cure* on the latent variable 'attitude towards multiple partnering' is -0.907. This means that knowledge that AIDS has no cure significantly decreases the value of the attitude of having multiple partners. The indirect effect on the indicators can be interpreted as follows: Multiplying this value by the factor loading for y_1 (i.e. 0.246) gives a value of -0.223. This implies that knowledge that AIDS has no cure has a significant negative effect, associated with a considerably decreased average number of girl friend partners. Now for y_2, multiplying the factor loading by the value of fixed effects for *knowledge that AIDS has no cure* gives a value of -1.661. Similarly the result for y_3 is -1.341. This suggests that, knowledge that AIDS has no cure is significantly associated with decreased average number of girl friends, number of casual, and number of commercial partners. The direction of significance for knowledge of symptoms of STIs, religion denomination, level of education attained by the respondents, knowledge that a healthy looking person can be HIV positive and length of stay at the place of survey is also similar to that of knowledge that AIDS has no cure. For instance, considering the effect of religion, the effect of *Christianity* on the latent variable is -1.468. Multiplying this value by the factor loading for y_1 (i.e. 0.246) gives a value of -0.361. This implies that being a Christian has a significant negative effect, associated with a decreased average number of girl friend partners compared to the reference category none/traditional. Multiplying the factor loadings by the value of fixed effects for *Christianity* gives an effect of -2.688 for y_2. Similarly the result for y_3 is -2.171. This suggests that, Christianity is significantly associated with decreased average number of girl friends, number of casual, and number of commercial partners. The direction of significance for Islam is similar to Christianity but with higher absolute magnitudes. This also suggests that Islam is significantly associated with decreased average number of girl friends, number of casual, and number of commercial partners, compared to the reference category none/traditional. Considering the effect of level of educational attainment, all three levels have about the same significant effect of reducing the attitude of having multiple partners.

Although the effect of locality (rural/urban) where respondent was during the time of the survey was positively associated with attitudes towards multiple partnering, this effect was not significant. On the other hand, knowledge of modes of transmission and modes of prevention are positively and significantly associated with increased average number of girl friend, casual and commercial sex partners. For instance, the effect of knowledge of mode of transmission is 1.196, multiplying this by the respective factor loading gives 0.294 for y_1, 2.190 for y_2 and 1.769 for y_3, while multiplying the effect (0.636) of knowledge of mode of prevention gives 0.157 for y_1, 1.165 for y_2 and 0.941 for y_3. These findings reveal that effects of knowledge about modes of transmission and modes of prevention are positively and significantly associated with increased average number of girl friends, number of casual, and number of commercial partners. This calls for intensive and modified interventions as knowledge about modes of transmission and modes of prevention does not translate into adopting a safer sex practice especially in the context of partner reduction. These findings are similar to what has been reported in literature. For instance, see FMOH [Nigeria] (2003); FMOH [Nigeria] (2006b) and FMOH [Nigeria] (2008).

From Table 3 we get the following effects of marital status. The intercept terms 1.190, 1.739 and 2.610 represent the effects of the reference category 'never married' on y_1, y_2, and y_3. The effects of 'formerly married' and 'currently married' are obtained by adding the respective estimates of direct effects to the intercept terms in Table 3. (Note, however, that the effects in Table 3 are not significant for 'formerly married'). For example, the effects of 'currently married' are obtained as -0.725, 0.446 and 1.435 respectively. This implies that effect of being currently married is significantly associated with decreased average number of girl friends but is positively and significantly associated with increased average number of casual and commercial partners. This again has pragmatic implications as respondents that are currently married are associated with increased number of casual and commercial partners which results into trans-generational and concurrent partners. This increases the likelihood of contracting HIV.

Turning attention to the non-linear effects of the continuous covariates in the data, Figure 2 provides findings about the non-linear effects of the respondents' age and age at first sex. The first panel shows an approximately quadratic pattern for respondents' age. Evidently effect of age is non-linear, and an assumption of linear effect would have resulted in erroneous and spurious conclusions. Figure 2 shows that there is a considerably increased attitude for having multiple partners up to approximately age 28. A noticeable steady decrease in average number of girl friends is evident beyond age 30. In a similar manner, effect of respondents' age at first sex has an approximately 'U' shaped effect on the attitude of having multiple partners.

Results of the spatial effects for the fitted model are displayed in Figure 3 (*a* and *b*). Figure 3 shows that there exists substantial geographical variations in attitude towards multiple sexual partnering across Nigeria. While some states were significantly associated with increased number of sexual partners, some were significantly associated with decreased number of sexual partners. Figure 3b presents the map of significance of the spatial effects. States in white colour have positive credible intervals, states in black have negative credible intervals, while states in grey colour have credible intervals that include 0 (zero). Adamawa, Kaduna, Ogun and Ondo states are significantly associated with increased number of girl friends, casual and commercial partners while Anambra, Delta, Edo, FCT, Gombe, Jigawa and Kano were associated with decreased numbers of sexual partners after controlling for other covariates.

5. Discussion of Results and Conclusions

In sub-Saharan African and other developing countries, many men (and women) have multiple concurrent sexual partners. Sexual relationships with more than one partner greatly increase the risk of exposure to HIV, and the practise of multiple partner relationships is often highlighted as a major factor in the spread of HIV in those regions. Unprotected sex with casual and commercial partners highly increases the risk of contracting HIV and other sexually transmitted diseases. Studies have identified multiple concurrent partnerships, including long relationships, as a driving force of HIV epidemic. In this paper, we propose the use of geoadditive latent variable modelling of count data for exploring possible association between number of sexual partners and possible covariates in Nigeria. This method of analysis flexibly models the relationship by jointly adjusting for possible geographical variations, non-linear effects, direct and indirect effects as well as the factors loadings for the indicators. To the best of our knowledge, there has been no previous work where a latent variable model of count data has been considered especially with incorporation of a spatial component in a generalized additive concept.

This study was conducted with the aim of providing policy makers with tools to enhance the design of appropriate effective HIV prevention strategies. Previous studies have shown that respondent's age can be a risk factor for multiple sexual partners. In this paper, respondent's age is shown to be significantly related to the number and types of sexual partners. Evidently, this relationship is non-linear (see Figure 2). Furthermore, as found in other studies, age at first coitus was significantly associated with number of sexual partners. Santelli et al. (1998) found that having been young at first coitus is a risk factor for having more than one sexual partner. In our study, effect of age at first sex was nonlinearly related to multiple sexual partners. This effect is approximately symmetrical around age 28 years for different types of sexual partners. Respondents who had their first sexual experience at age below 25 years are more likely to have multiple girl friends, casual and commercial partners.

Considering the level of casual (62.1%) and commercial (62.3%) sexual activities within the last 12 months prior to the survey in Nigeria among the respondents, HIV programmers, government and other stakeholders need to come together to design an effective HIV prevention programme to reduce multiple concurrent sexual partnerships. Findings from this paper reveal religion differentials in attitude about multiple partnering. While, officially the doctrine of Islam permits a man to marry four wives, it however, discourages non-marital sexual practice. This is not to say that Christians do not engage in polygamy, however, it is less prevalent among Christians.

Spatial effects obviously reveal that multiple partnering varies according to geographical locations, i.e. states in Nigeria. Formerly married respondents are more likely to have increased average number of girl friends and casual partners.

In conclusion, findings from this paper provide insight to policy formulation. Scarce resources have been identified as a major challenge towards implementation of necessary intervention strategies in sub-Saharan African countries, including Nigeria. This paper provides policy-makers with tools to enhance appropriate policy formulation on the prevention of HIV/AIDS; which can also assist in allocating resources to states or districts where the resources can be effectively utilized. While identifying states that require intensive prevention efforts towards the partner reduction, the need for sustenance of the low number of sexual partners in states that are associated with low number of sexual partners must be ensured by policy-makers in the affected states.

6. Supplementary materials: No supplementary materials.

Acknowledgements

The authors would like to appreciate the Federal Ministry of Health Nigeria for releasing this data for this paper. We thank Iris Burger for the preparation of this manuscript.

References

Adetunji, J. and Meekers, D. (2001). Consistency in Condom Use in the Context of HIV/AIDS in Zimbabwe. *Journal of Biosocial Science* **33**, 121-138.

Brezger, A. and Lang, S. (2006). Generalized Structured Additive Regression based on Bayesian P-splines. *Computational Statistics and Data Analysis* **50**, 967-991.

Eilers, P.H.C. and Marx, D.B. (1996). Flexible Smoothing with B-Splines and Penalties. *Statistical Science* **11(2)**, 89-121.

Fahrmeir, L. and Raach, A.W. (2007). A Bayesian Semiparametric Latent Variable Model for Mixed Responses. *Psychometrica* **72(3)**, 327-346.

Fahrmeir, L. and Steinert, S. (2006). A geoadditive Bayesian latent variable model for Poisson indicators. *Sonderforschungsbereich 386, Discussion Paper* 508, Ludwig Maximilian University of Munich, Germany.

Federal Ministry of Health [Nigeria] (2003). National HIV/AIDS & Reproductive Health Survey (NARHS), 2003. Federal Ministry of Health, Abuja, Nigeria.

Federal Ministry of Health [Nigeria] (2006a): *The 2005 National HIV/Syphilis Sero-prevalence Sentinel Survey among Pregnant Women Attending Antenatal Clinics in Nigeria.* Federal Ministry of Health, Abuja, Nigeria.

Federal Ministry of Health [Nigeria] (2006b). National HIV/AIDS & Reproductive Health Survey (NARHS), 2005. Federal Ministry of Health, Abuja, Nigeria.

Federal Ministry of Health [Nigeria] (2008a). *HIV/STI Integrated Biological and Behavioural Surveillance Survey (IBBSS).* Federal Ministry of Health, Abuja, Nigeria.

Federal Ministry of Health [Nigeria] (2008b). National HIV/AIDS & Reproductive Health Survey (NARHS), 2005. Federal Ministry of Health, Abuja, Nigeria.

Flood, M. (2003). Lust, Trust and Latex: Why Young Heterosexual Men DO not Use Condoms. *Culture Health Sexuality* **5**, 353-369.

Fruehwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary Mixture Sampling for parameter-driven Models of time Series of Counts with Application to State Space Modelling. *Biometrika* **93(4)**, 827-841; doi:10.1093/biomet/93.4.827.

Hearst, N. and Chen, S. (2004). Condom Promotion for AIDS Prevention in the Developing World: Is it working? *Studies in Family Planning* **35**, 39-47.

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

Lopes, H.F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14,** 41–67.

Meekers, D., Klein, M., and Foyet, L. (2003). Patterns of HIV Risk Behaviour and Condom Use among Youth in Yaounde and Douala, Cameroon. *AIDS Behaviour* **7**, 413-420.

Morris, M. (2002). A Comparative Study of Concurrent Sexual Partnerships in the United States, Thailand and Uganda. *American Sociology Association Annual Meeting Published Abstracts, Anaheim California, Aug 18-21: Session 409.*

Quinn, K.M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* **12**, 338–353.

Raach, A.W. (2005). A Bayesian Semiparametric Latent Variable Model for Binary, Ordinal and Continuous Response. Dissertation, available from http://www.educ.ub.ub.uni-muenchen.de.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall.

Sammel, M.D., Ryan, L.M., & Legler, J.M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society Series B*, **59**, 667–678.

Santelli, J.S., Brener, N.D., Lowry, R., Bhatt, A., and Zabin, L.S. (1998). Multiple Sexual Partners among U.S. Adolescents and Young Adults. *Family Planning Perspectives* **30(6)**, 271-275.

Appendix

Inference is based on a data augmentation approach for auxiliary Gaussian responses. This facilitates full Bayesian inference via Gibbs sampling, and it allows us to combine geoadditive latent variable models developed in Fahrmeir and Raach (2007) for binary, ordinal and continuous indicators with models for count indicators considered here. Following a recent suggestion of Frühwirth-Schnatter and Wagner (2006) in the context of state space models for count data, the introduction of two so called data augmentation steps eliminates the nonlinearity of the Poisson model as well as the non-normality of the error term. The (conditional) distribution of $y_{ij} \mid \mu_{ij}, \cdot$ is considered as the distribution of the number of jumps of an unobserved Poisson process in the time interval [0, 1]. The first data augmentation step introduces the inter-arrival times τ_{ijl}, i = 1,…, n, j = 1,…, p, l = 1,…, y_{ij} + 1, of this unobserved Poisson process. They follow an exponential distribution $\tau_{ijl} \sim \text{Exp}(\mu_{ij}) = \text{Exp}(1)/\mu_{ij}$, leading to the (log)-linear model

$$-\log \tau_{ijl} = \alpha_j' u_i + \lambda_j v_i + \varepsilon_{ijl}, \quad \varepsilon_{ijl} \sim \log \text{Exp}(1).$$

Following Frühwirth-Schnatter and Wagner (2006), we approximate the density of the logExp(1)-distribution of the error term ε_{ijl} by a mixture of ten normal distributions to obtain a conditionally Gaussian model

$$f(\varepsilon_{ijl}) \approx \sum_{r=1}^{10} w_r f_N(\varepsilon_{ijl}, m_r, \sigma_r^2)$$

where weights w_r, means m_r and variances σ_r^2 are calculated by minimizing the Kullback-Leibler distance, see Frühwirth-Schnatter and Wagner (2006, Table 1).

This way we obtain a linear model for log-interarrival times, with errors following a Gaussian mixture distribution with known weights, mean and variances. For Gibbs sampling, a second data augmentation step introduces a vector c of latent indicators for the components of the Gaussian mixture distributions following a multinomial distribution, with class probabilities given by the weights. Conditional on mixture component indicators, we finally obtain additional $y_{ij}+1$ Gaussian models for a $(y_{ij}+1)$-dimensional vector y*, with means m_{r} as an additional offset, see Fahrmeir and Steinert (2006) for details.

Full Bayesian inference can be carried out via Gibbs sampling in combination with data augmentation, considering underlying Gaussian variables y* and latent variables *v* as additional "parameters". The first step of the Gibbs sampler generates interarrival times, component indicators c and auxiliary Gaussian variables y*. The remaining steps, generating latent variables v and drawing from full conditionals for all parameters, are essentially the same as in Fahrmeir and Raach (2007) and Raach (2005). Further details are given in Fahrmeir and Steinert (2006).

Table 1: Description of variables used in the models

Variable	Description	Scales of measurement
Outcome variable Multiple sexual partnering	• Number of spousal/cohabiting partners • Number of boy/girl friends • Number of casual partners • Number of commercial partners	Counts (this is obtained as a composite index that sums the total number of different sexual partners)
Independent variables		
- Current age of respondents	- Age of respondents (measured in years) as at the last birthday	Continuous
- Age at first sex	- Reported age (measured in years) at which respondents had first sexual experience	Continuous
- Length of stay	- Duration of stay in the place of residence. This was measured in years	Continuous
- Educational attainment	- Highest level of education attained by the respondents	Categorical: No formal education (reference), primary, secondary or higher
- Religion	- The religion affiliation a respondent belongs to was incorporated into the survey	None/Traditional (reference), Islam, Christianity
- Locality of residence	- Place of residence of residence was meant at distinguishing those who resided in rural from those who resided in urban areas as at the time of the survey	Rural or Urban (reference)
- Being away from home for more than one month in the last one year	- This is a variable that measures whether a respondent has been away from home for more than 30 days during the last 12 months	Dichotomous: Yes or No
- Knowledge of symptoms of STIs	- Correct knowledge of symptoms of STIs in men and women: painful urination, genital discharge, genital sore/ulcer	Dichotomous: Yes (all three) or No (otherwise: reference)
- States	- This is an administrative boundary in Nigeria	
- Healthy looking person	- Knowledge that a healthy looking person can be HIV positive	Dichotomous: Yes vs. No
- AIDS has no cure	- Knowledge that AIDS has no cure	Dichotomous: Yes vs. No
- Correct mode of transmission[1]	- Correct knowledge about modes of transmission of HIV: sexual intercourse, blood transfusion, transmission from mother to an unborn child, sharing of sharp objects like razors and needles	Dichotomous: a composite index that sums all affirmative responses to these questions (1 if yes to all, else 0).
- Correct mode of prevention	- Knowledge about modes of prevention of HIV: staying with one uninfected partner, using condom every time, abstaining from sex, reducing the numbers of sexual partners, avoid sharing of sharp objects like needles and razors, avoid sex with commercial sex workers, delaying the onset of sexual intercourse, avoid having sex with people who have many sexual partners	Dichotomous: a composite index that sums all affirmative responses to these questions (1 if yes to all, else 0).
- Marital status	- Respondent's marital status	Categorical: Never married (reference), currently married/cohabiting with a sexual partner, formerly married (widowed, separated or divorced)

[1] Composite UNAIDS indicator for mode of transmission was used in this paper, i.e. respondents who know that HIV can be transmitted through sexual intercourse, blood transfusion, transmission from mother to an unborn child, sharing of sharp objects like razors and needles

Table 2: Posterior means with 95% credible intervals for factor loadings

Variable	Posterior mean	Std. dev (error)	95% Credible Interval Lower	Upper
λ_{11}	0.246	0.033	0.186	0.311
λ_{21}	1.831	0.343	1.027	2.191
λ_{31}	1.479	0.170	1.150	1.740

Table 3: Posterior means with 95% credible intervals for direct effects of marital status

Variables	Posterior mean	Std. dev (error)	95% Credible Interval Lower	Upper
Never married				
a_{10}	1.190	0.146	0.910	1.476
a_{20}	1.739	1.033	-0.222	3.185
a_{30}	2.610	0.794	0.917	3.700
Formerly married				
a_{11}	0.115	0.294	-0.488	0.671
a_{21}	-7.450	6.899	-23.772	0.210
a_{31}	1.490	0.910	-0.534	3.120
Currently married				
a_{12}	-1.915	0.112	-2.136	-1.695
a_{22}	-1.293	0.492	-2.335	-0.503
a_{32}	-1.175	0.333	-1.945	-0.518

Table 4: Posterior means with 95% credible intervals for indirect effects

Variable	Posterior mean	Std. dev (error)	95% Credible Interval Lower	Upper
Away from home	0.071	0.154	-0.229	0.367
Knowledge of symptoms of STIs	-0.788	0.284	-1.317	-0.232
Knowledge that AIDS has no cure	-0.907	0.243	-1.415	-0.458
Rural	0.029	0.180	-0.309	0.372
Christianity	-1.468	0.338	-2.149	-0.799
Islam	-2.438	0.389	-3.257	-1.729
Primary	-0.955	0.382	-1.610	-0.110
Secondary	-1.055	0.387	-1.814	-0.304
Higher	-0.906	0.399	-1.649	-0.086
Knowledge of mode of transmission	1.196	0.471	0.266	2.099
Knowledge of mode of prevention	0.636	0.209	0.238	1.064
Knowledge that a healthy looking person can be HIV positive	-0.841	0.232	-1.302	-0.408
Length of stay	-0.026	0.007	-0.039	-0.013

Figure 1: Map of Nigeria showing geographical locations of states

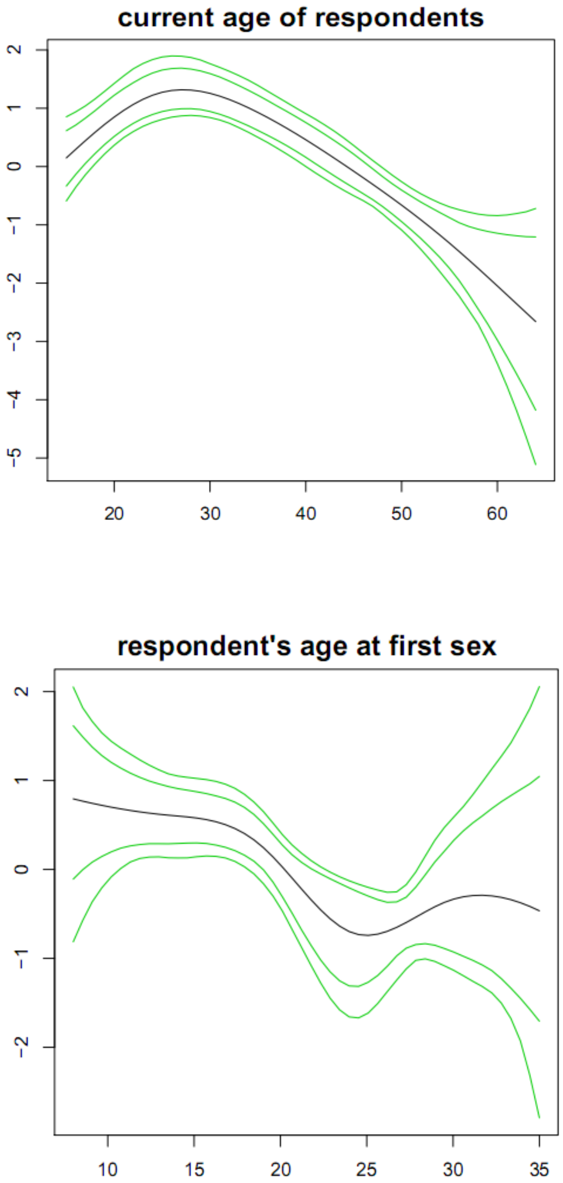

Figure 2: Non-linear effects of respondent’s age and age at first sexual intercourse.

a) Spatial effects

b) Map of significance of spatial effects

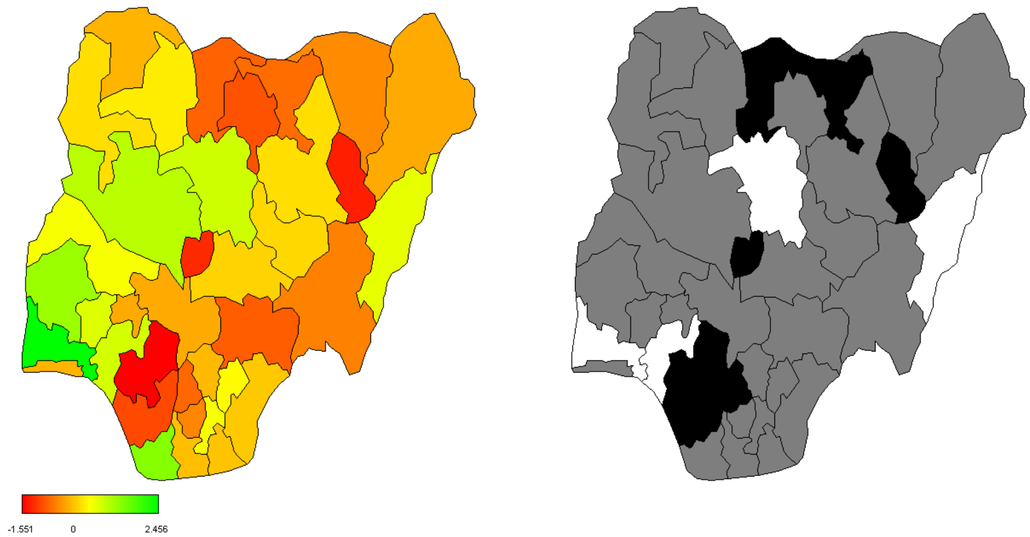

Figure 3: Spatial effects of attitude towards multiple sexual partnering in Nigeria. Shown are the posterior means for the states (a) and 95% point-wise credible intervals (b) for the fitted model.