Journal of Applied Statistics, 2015
http://dx.doi.org/10.1080/02664763.2015.1022138

Taylor & Francis
Taylor & Francis Group

Joint modeling of correlated binary outcomes: HIV-1 and HSV-2 co-infection

Musie Ghebremichael*

Harvard Medical School and Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA

(*Received 4 September 2014; accepted 19 February 2015*)

Herpes Simplex Virus Type 2 (HSV-2) facilitates the sexual acquisition and transmission of HIV-1 infection and is highly prevalent in most regions experiencing severe HIV epidemics. In sub-Saharan Africa, where HIV infection is a public health burden, the prevalence of HSV-2 is substantially high. The high prevalence of HSV-2 and the association between HSV-2 infection and HIV-1 acquisition could play a significant role in the spread of HIV-1 in the region. The objective of our study was to identify risk factors for HSV-2 and HIV-1 infections among men in sub-Saharan Africa. We used a joint response model that accommodates the interdependence between the two infections in assessing their risk factors. Simulation studies show superiority of the joint response model compared to the traditional models which ignore the dependence between the two infections. We found higher odds of having HSV-2/HIV-1 among older men, in men who had multiple sexual partners, abused alcohol, or reported symptoms of sexually transmitted infections. These findings suggest that interventions that identify and control the risk factors of the two infections should be part of HIV-1 prevention programs in sub-Saharan Africa where antiretroviral therapy is not readily available.

Keywords: AIDS; HIV-1; HSV-2; bivariate binary outcomes; joint modeling; Sub-Saharan Africa

1. Introduction

HIV continues to spread around the world. Based on the latest UNAIDS report on the global AIDS epidemic [32], an estimated 35.3 million people worldwide are living with HIV. Sub-Saharan Africa remains the region most heavily affected by HIV. Seventy percent of all people living with HIV/AIDS are found in sub-Saharan Africa [32]. Moreover, sub-Saharan Africa has the highest prevalence of other sexually transmitted infections (STIs) [22]. There is compelling evidence that STIs facilitate the acquisition and transmission of HIV infection [9,11]. Thus, identification and treatment of STIs as well as interventions that control the risk factors associated with acquisition of STIs are essential for successful HIV prevention in sub-Saharan Africa. In the absence of an effective vaccine and universal access to antiretroviral therapy, STIs control programs should become part of the HIV prevention and care package in the region.

*Email: musie_ghebremichael@dfci.harvard.edu

Herpes Simplex Virus Type 2 (HSV-2) infection is the most common STI in the world. It causes most of the genital herpes cases and is transmitted through forms of sexual contact with an infected person who is shedding the virus. Most HIV-1 infected persons are co-infected with HSV-2 and as a result, HSV-2 is highly prevalent in regions experiencing severe HIV epidemics [6,12]. In sub-Saharan Africa, where HIV-1 prevalence is the highest, substantially higher HSV-2 rates are seen with prevalence in adults of around 80% [28]. The proportion of HIV-1 infections attributable to HSV-2 infection in the region was estimated to be 25–35 % and has increased over time with maturity of the HIV epidemic [1,10]. HSV-2 plays a major role in facilitating HIV-1 acquisition and transmission. It creates potential biological bridges for HIV acquisition, either directly by eroding the genital epithelial barrier making a path through which HIV-1 can enter, or indirectly by causing genital inflammation which leads to the recruitment of new HIV-1 targets cells such as activated CD4 T cells and macrophages that reside in the sub epithelium [7,17]. Moreover, in persons infected with HIV-1, plasma and genital HIV-1 concentrations increase during reactivation of HSV-2 [27,28], thereby enhancing HIV-1 transmission and progression. This suggests that treatment of HIV-1 infected patients with HSV-2 drugs which are inexpensive and easy to administer might be an appealing strategy to fight AIDS in sub-Saharan Africa and other resource-poor countries where HIV-1 treatment is not readily available. However, the role that HSV-2 suppressive therapy plays in HIV control is still controversial. Several efficacy trials of HSV-2 suppression for prevention of HIV resulted in mixed findings [21,25].

The high prevalence of HSV-2 and the association between HSV-2 infection and HIV-1 acquisition could play a major role in the spread of HIV-1 in sub-Saharan Africa. Promoting awareness of HSV-2, screening for HSV-2, treating HSV-2 infection, and identifying the shared risk factors of the two infections may thus play important roles in the prevention of HIV-1 infection in the region. The objective of the present study was to identify potential risk factors for HSV-2 and HIV-1 infections among men in sub-Saharan Africa. In the region, sexual decision-making is the prerogative of men. Thus, interventions that target risky behaviors in men are vital for HIV-1 prevention in the region. We used a joint response model that accommodates the interdependence between HSV-2 and HIV-1 infections in assessing the socio-demographic, behavioral and biological risk factors of the two infections. Separate analysis of each infection ignores the dependence between the two infections and analyzing the data as if it were independent results in biased estimates. Joint modeling of the two outcomes provides better control over the type I error rates in multiple tests and gains in efficiency in the parameter estimates. Previous works on joint modeling of correlated outcomes have largely focused on continuous outcomes or a mixture of continuous and discrete outcomes [14,24,30]. In the present study, however, we modeled the bivariate response of correlated binary outcomes as functions of explanatory variables while simultaneously taking into consideration the correlation between the two outcomes.

The remainder of the paper is organized as follows: In Section 2, we introduce the statistical models used: joint response and marginal models. Furthermore, in Section 2, we compare the regression parameters from the two models. Simulation studies were performed to determine if the joint response model has increased efficiency relative to the marginal models. Results of these simulation studies are summarized in Section 3. In Section 4, we introduce the study population, study design and study variables. Furthermore, in Section 4 we summarize the results of our data analyses. Finally, Section 5 features interpretations of our findings and concluding remarks.

2. Statistical methods

2.1 *Marginal and joint models*

Let Y_{ij} denote the ith response ($i = 1, 2$) of the jth ($j = 1, 2, \ldots, n$) subject with $i = 1$ for HSV-2 and $i = 2$ for HIV-1. Y_{ij} is a binary random variable and takes the values 0 or 1 depending on

whether response occurs or not. Thus, for the jth subject, we have a bivariate binary response vector (Y_{1j}, Y_{2j}). Let X_{1j} and X_{2j} be vectors of covariates associated with HSV-2 and HIV-1, respectively, and β_1 ($\tilde{\beta}_1$), and β_2 ($\tilde{\beta}_2$) are the corresponding vectors of regression coefficients. In the sequel, dependence on the covariates will be suppressed from the notations. The association between the covariates and each infection can be examined using separate regression models for each outcome:

$$\begin{aligned}\text{logit}\{\text{E}(Y_{1j})\} &= \text{logit}\{\Pr(Y_{1j} = 1|X_{1j}, \beta_1)\} \\ &= \beta_1^{\text{T}} X_{1j} \end{aligned} \tag{1}$$

and

$$\begin{aligned}\text{logit}\{\text{E}(Y_{2j})\} &= \text{logit}\{\Pr(Y_{2j} = 1|X_{2j}, \beta_2)\} \\ &= \beta_2^{\text{T}} X_{2j}. \end{aligned} \tag{2}$$

These traditional logistic regression models ignore the correlation between the two infections. However, outcomes from the same individual are correlated since HIV-1 infected patients are more likely to be infected with HSV-2, and vice versa. We assume a set of latent, unobserved, random effects are shared by the two infections of the same individual, that is, the two infections share common unobservable features. We therefore account for the correlation between the two infections of the same individual using a shared random intercept. The random intercept captures the unobserved factors specific to each individual which may influence the responses. More specifically, let b_j denote the random intercept shared by the two infections of the jth ($j = 1, 2, \ldots, n$) individual. Let d_{1i} and d_{2i} define dummy variables with $d_{1i} = 1$ for $i = 1$ and $d_{2i} = 1$ for $i = 2$. Then, the joint response model is given by

$$\begin{aligned}\text{logit}\{\text{E}(Y_{ij}|b_j)\} &= \text{logit}\{\Pr(Y_{ij} = 1|X_{ij}, \tilde{\beta}_i, b_j)\} \\ &= d_{1j}(\tilde{\beta_1}^{\text{T}} X_{1j} + b_j) + d_{2j}(\tilde{\beta_2}^{\text{T}} X_{2j} + b_j). \end{aligned} \tag{3}$$

In this equation, the bivariate responses (Y_{1j}, Y_{2j}) of all individuals are stacked into a single response vector. The random intercepts are assumed to vary independently from one individual to another. In addition, the random intercepts are assumed to be normally distributed with zero mean and variance σ_b^2. Let G denote the distribution of b_j. The joint responses of an individual are assumed to be independent given the shared random intercept. By exploiting the conditional independence from this assumption, we can write the likelihood function of the joint response model as follows:

$$\begin{aligned}\text{L}(\cdot) &= \prod_{j=1}^{n} \Pr(Y_{ij} = 1|X_{ij}, \tilde{\beta}_i, b_j) \\ &= \prod_{j=1}^{n} \left\{ \int \Pr(Y_{ij} = 1|X_{ij}, \tilde{\beta}_i, b_j)\, \text{d}G(b_j) \right\} \\ &= \prod_{j=1}^{n} \left\{ \int \prod_{i=1}^{2} \Pr(Y_{ij} = 1|X_{ij}, \tilde{\beta}_i, b_j)\, \text{d}G(b_j) \right\} \\ &= \prod_{j=1}^{n} \left\{ \int \prod_{i=1}^{2} \frac{\text{e}^{b_j + \tilde{\beta}_i^{\text{T}} X_{ij}}}{1 + \text{e}^{b_j + \tilde{\beta}_i^{\text{T}} X_{ij}}}\, \text{d}G(b_j) \right\}. \end{aligned} \tag{4}$$

Maximum likelihood estimation was used to obtain estimates of model parameters ($\tilde{\beta}_1$, $\tilde{\beta}_2$, σ_b^2). The estimation was based on the maximization of the log-likelihood function. Note that maximizing the likelihood function is equivalent to maximizing the logarithm of the likelihood function. The integrals involved in Equation (4) do not have closed-form solutions, and hence the maximization of the log-likelihood function required approximations. Therefore, integral approximation through Gaussian quadrature was used to maximize the joint log likelihood. More specifically, parameters of the joint model were estimated simultaneously via maximum likelihood by evaluating the integrals using Gaussian adaptive quadrature approximation [8,29]. Adaptive quadrature provides a good approximation of the integral, especially when used for ordinal data with small sample sizes, and it is more feasible than classical numerical techniques in the presence of random effects [3,26]. Furthermore, adaptive quadrature-based maximum likelihood estimators have desirable asymptotic properties. In particular, maximum likelihood estimators based on adaptive quadrature are consistent and asymptotically normally distributed [2,20].

2.2 *Comparison of model parameters*

The regression parameters in the joint response and separate models do not have the same interpretations. The parameters in the joint response model have to be interpreted conditional on the random intercepts as opposed to the unconditional interpretations of the parameters in the separate models. The former are subject-specific (conditional) parameters and the latter are population-averaged (marginal) parameters. Thus, the two sets of parameters are not directly comparable. Marginal effects of the covariates in Equation (3) are required to compare the parameters in the joint response model to those obtained from the separate models. The subject-specific parameters of the joint model can be marginalized as described below.

From the separate models (1)–(2), we can write the marginal mean as $\mathrm{E}(Y_{ij}) = e^{\beta_i^{\mathrm{T}} X_{ij}}/(1 + e^{\beta_i^{\mathrm{T}} X_{ij}})$, for $i = 1, 2$. Similarly, for the ith response in the joint response model (3), the conditional mean is given by $\mathrm{E}(Y_{ij}|b_j) = e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}/(1 + e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}})$. The marginal information can be obtained from the joint response model by taking the expectation over the random intercepts as shown below:

$$
\begin{aligned}
\mathrm{E}(Y_{ij}) &= \mathrm{E}\{\mathrm{E}(Y_{ij}|b_j)\} \\
&= \mathrm{E}\left\{ \frac{e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}}{1 + e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}} \right\} \\
&= \int_{-\infty}^{\infty} \frac{e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}}{1 + e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}} \, \mathrm{d}G(b_j) \\
&= \int_{-\infty}^{\infty} \frac{e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}}{1 + e^{b_j + \tilde{\beta}_i^{\mathrm{T}} X_{ij}}} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-b_j^2/2\sigma_b^2} \, \mathrm{d}b_j \\
&\neq \frac{e^{\beta_i^{\mathrm{T}} X_{ij}}}{1 + e^{\beta_i^{\mathrm{T}} X_{ij}}}.
\end{aligned}
$$

There is no exact closed-form solution for the marginal mean in the above equation and hence we used approximation of the marginal mean as in [33]. That is, we approximated the logit of the marginal mean by $\mathrm{logit}\{\mathrm{E}(Y_{ij})\} \approx \tilde{\beta}_i^{\mathrm{T}} X_{ij}/\sqrt{1 + (16\sqrt{3}/15\pi)^2 \sigma_b^2}$. We used this approximation to

convert the subject-specific regression parameters of the joint response model to population-averaged parameters:

$$\tilde{\beta}_i^{*T} \approx \frac{1}{\sqrt{1+(16\sqrt{3}/15\pi)^2\sigma_b^2}}\tilde{\beta}_i^{\mathrm{T}}, \tag{5}$$

where $\tilde{\beta}_i^{*}$ $(i = 1, 2)$ are population-averaged parameters which correspond to the subject-specific parameters ($\tilde{\beta}_i$, $i = 1, 2$) of Equation (3). The relationship in Equation (5) indicates that the marginal parameters are smaller or equivalent to the conditional parameters. The discrepancies between the two parameters are dependent on the magnitude of the between subjects variation. The marginal parameters are approximately equivalent to the conditional parameters when all individuals share the same random intercept value (when there is no between subjects variation, i.e. $\sigma_b^2 = 0$). However, when $\sigma_b^2 > 0$, the marginal parameters are smaller in absolute value than the conditional parameters. The greater the between subjects variation, the smaller the marginal parameters are compared to the conditional parameters.

We now have both population-level and individual-level information available in a single model, that is, we have both marginal ($\tilde{\beta}_i^{*}$) and conditional ($\tilde{\beta}_i$) information from the joint response model. The marginal parameters are comparable to the marginal parameters obtained by modeling each outcome independently. We used the delta method to obtain approximate standard errors for the $\tilde{\beta}_i^{*}$'s when comparing the two sets of marginal parameters.

3. Simulation studies

Simulation studies were performed to examine the performance of the joint response model compared to the marginal models. In each simulation, we generated correlated binary outcomes (Y_{1j}, Y_{2j}), a binary covariate (X_{1j}) and a continuous (X_{2j}) covariate. The correlated binary outcomes were generated using the methods of Lipsitz *et al.* [19]. The covariates X_{1j} and X_{2j} were generated from Bernoulli (mean = 0.5) and normal (mean = 30, variance = 25) distributions. As defined in Section 2, Y_{ij} $(i = 1, 2;\ j = 1, 2, \ldots, n)$ is a binary random variable and takes the values 0 or 1 (negative or positive response). Let p_1 (p_2) denote the proportion of individuals on the first (second) outcome with a positive response and let ρ denote the correlation between the two outcomes. Three sets of simulation studies were considered. In the first part of our simulation study, we investigated the effect of sample size (n) on the performance of

Table 1. Parameter and standard error estimates from simulation study with varying sample size.

	Sample size			
Parameter	$n = 150$	$n = 200$	$n = 250$	$n = 300$
Joint				
$\tilde{\beta}_{11}^{*}$	0.0138 (0.4993)	0.0249 (0.4213)	− 0.0083 (0.3735)	− 0.0019 (0.3402)
$\tilde{\beta}_{12}^{*}$	− 0.0005 (0.0494)	− 0.0015 (0.0425)	0.0016 (0.0375)	0.0036 (0.0342)
$\tilde{\beta}_{21}^{*}$	0.0120 (0.4303)	0.0085 (0.3725)	− 0.0009 (0.3304)	− 0.0070 (0.3009)
$\tilde{\beta}_{22}^{*}$	0.0001 (0.0435)	− 0.0010 (0.0375)	0.0008 (0.0333)	0.0026 (0.0303)
Y_1				
β_{11}	0.0381 (0.5869)	0.0306 (0.4975)	− 0.0090 (0.4386)	− 0.0041 (0.3980)
β_{12}	− 0.0009 (0.0575)	− 0.0014 (0.0491)	0.0023 (0.0435)	0.0040 (0.0396)
Y_2				
β_{21}	0.0179 (0.4800)	0.0078 (0.4103)	− 0.0018 (0.3641)	− 0.0099 (0.3292)
β_{22}	0.0001 (0.0477)	− 0.0005 (0.0409)	0.0010 (0.0364)	0.0028 (0.0329)

the joint response model. We considered various values of n. For each n, we generated 1000 independent data sets with the two binary outcomes sharing the same covariates. From Equations (1) and (2), the marginal models in our simulation study can be written as $\text{logit}\{\text{E}(Y_{1j})\} = \beta_{10} + \beta_{11}X_{1j} + \beta_{12}X_{2j}$ and $\text{logit}\{\text{E}(Y_{2j})\} = \beta_{20} + \beta_{21}X_{1j} + \beta_{22}X_{2j}$. Similarly, from Equation (3), the joint response model can be written as $\text{logit}\{\text{E}(Y_{ij}|b_j)\} = d_{1j}(\tilde{\beta}_{10} + \tilde{\beta}_{11}X_{1j} + \tilde{\beta}_{12}X_{2j} + b_j) + d_{2j}(\tilde{\beta}_{20} + \tilde{\beta}_{21}X_{1j} + \tilde{\beta}_{22}X_{2j} + b_j)$. These models were applied to the generated 1000 independent data sets, and the means of the 1000 different parameter estimates together with their corresponding standard errors, were calculated. Marginal parameters of the joint response model were obtained using the approximate relationship in Equation (5) and compared with marginal parameters obtained by modeling each outcome independently. In the second part of our simulation study, our objective was to investigate the performance of the joint response model for various values of ρ. We considered ρ values of 0, 0.25, 0.50 and 0.70. For each ρ, we conducted our simulation study in the same format as in part 1. Not all participants in our study were tested for STIs thereby resulting in several subjects with unknown HIV-1 or HSV-2 status. Thus, we conducted

Table 2. Parameter and standard error estimates from simulation study with varying correlation.

	Correlation coefficient			
Parameter	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.70$
Joint				
$\tilde{\beta}^*_{11}$	0.0042 (0.4894)	0.0070 (0.4521)	0.0249 (0.4213)	0.0091 (0.3029)
$\tilde{\beta}^*_{12}$	0.0021 (0.0483)	− 0.0002 (0.0451)	− 0.0015 (0.0425)	0.0005 (0.0309)
$\tilde{\beta}^*_{21}$	0.0056 (0.4046)	− 0.0110 (0.3905)	0.0085 (0.3725)	0.0011 (0.2343)
$\tilde{\beta}^*_{22}$	0.0007 (0.0404)	0.0004 (0.0392)	− 0.0010 (0.0375)	− 0.0001 (0.0242)
Y_1				
β_{11}	0.0044 (0.4975)	0.0074 (0.4978)	0.0306 (0.4975)	0.0318 (0.4968)
β_{12}	0.0020 (0.0490)	− 0.0002 (0.0490)	− 0.0014 (0.0491)	0.0009 (0.0492)
Y_2				
β_{21}	0.0060 (0.4080)	− 0.0115 (0.4097)	0.0078 (0.4103)	0.0234 (0.4089)
β_{22}	0.0007 (0.0407)	0.0003 (0.0408)	− 0.0005 (0.0409)	− 0.0002 (0.0408)

Table 3. Parameter and standard error estimates from simulation study with missing data.

	% of Data missing			
Parameter	0%	10%	20%	30%
Joint				
$\tilde{\beta}^*_{11}$	0.0249 (0.4213)	− 0.0037 (0.4397)	− 0.0345 (0.4673)	− 0.0076 (0.4993)
$\tilde{\beta}^*_{12}$	− 0.0015 (0.0425)	0.0004 (0.0439)	0.0010 (0.0467)	− 0.0001 (0.0495)
$\tilde{\beta}^*_{21}$	0.0085 (0.3725)	0.0013 (0.3854)	− 0.0024 (0.4034)	− 0.0071 (0.4299)
$\tilde{\beta}^*_{22}$	− 0.0010 (0.0375)	0.0001 (0.0387)	0.0013 (0.0405)	− 0.0002 (0.0429)
Y_1				
β_{11}	0.0306 (0.4975)	− 0.0006 (0.5213)	− 0.0426(0.5630)	− 0.0093 (0.6055)
β_{12}	− 0.0014 (0.0491)	0.0009 (0.0514)	0.0015 (0.0553)	0.0002 (0.0593)
Y_2				
β_{21}	0.0078 (0.4103)	0.0060 (0.4328)	− 0.0009 (0.4604)	− 0.0068 (0.4976)
β_{22}	− 0.0005 (0.0409)	0.0001 (0.04306)	0.0016 (0.0459)	− 0.0006 (0.0494)

an additional simulation study to examine the performance of the joint response model in the presence of missing data. The third part of our simulation study was conducted in the same way as in parts 1–2 with 0%, 10%, 20% and 30% missing data in each outcome variable. In all cases, we considered various values of p_1 and p_2.

The results presented here are representative of the many simulations performed. Tables 1–3 show results for $p_1 = 0.10$ and $p_2 = 0.15$. In Tables 1 and 3, a ρ-value of 0.5 was used. In Tables 2 and 3, a sample size of 200 subjects was used. The average parameter estimates and the average standards errors presented in these tables were estimated based on the 1000 independent data sets. These results indicate that the joint response model provides parameter estimates with smaller standard errors. The gain in efficiency is noticeable for smaller sample sizes, higher correlations and larger proportions of missing data.

4. Data analyses

4.1 *Study participants and laboratory methods*

This study was based on analysis of data collected as part of the Moshi Infertility Survey. The survey was conducted from November 2002 to March 2003 in the Moshi urban district of Tanzania and involved a two-stage sampling. In the first stage of sampling, a number of households were randomly selected from the district. A total of 794 men, who were either partners or husbands of the women residing in selected households, were interviewed. Information was collected on socio-demographic characteristics and high-risk sexual behaviors. Blood and urine samples were drawn to test for HIV-1, HSV-2 and other STIs (syphilis, Chlamydia, *Trichomonas*, and *Mycoplasma genitalium* infections). HIV-1 infection was determined using HIV enzyme-linked immunosorbent assay (ELISA), and reactive samples were confirmed using Wellcozyme HIV-1 ELISA test. Western blot tests were used to confirm discordant ELISA test results. A type-specific ELISA test was used to detect HSV-2. Rapid Plasma Reagin card test, Treponema Pallidum Hemagglutination Assay and real-time multiplex polymerase chain reaction assay were used to diagnose the other STIs. The study was approved by the Harvard School of Public Health IRB (HSC Protocol #0108ACOM), University of Maryland IRB (Protocol #05-0031), Kilimanjaro Christian Medical Center Ethics Committee, Research and Ethical Clearance of the Tanzanian National Institute for Medical Research, the Centers for Disease Control and Prevention Institutional Review Board. Written informed consent for participation in the study was obtained from participants. A detailed description of the study design and characteristics of the study participants has been described in previously published materials [18].

4.2 *Study measures*

Socio-demographic characteristics (age in years, education), high-risk behaviors (age at first sex, number of partners, alcohol abuse and condom use) and biological/medical factors (symptoms of STIs and STIs other than HIV-1/HSV-2) were considered as covariates in our analyses. Alcohol abuse was measured by the CAGE score [23] which is defined as the sum of yes answers to the following four questions: (1) Have you ever felt you should cut down on drinking?, (2) Have people annoyed you by criticizing your drinking?, (3) Have you ever felt bad or guiltily about your drinking? and (4) Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover? The CAGE score ranged from 0 to 4. Alcohol abuse was defined as a CAGE score of 2–4 and alcohol non-abuse as a CAGE score of 0–1. An STI's symptom was defined as having at least one of the following symptoms: abdominal pain, abnormal genital discharge, foul smell in the genital area, excessive genital secretions, swellings in the genital area, itching in the genital area, burning pain on micturition, pain during intercourse and genital ulcers.

4.3 *Results*

Five hundred and sixty-seven out of the 794 (71.4%) men enrolled in the study provided blood samples for HIV-1 and HSV-2 infections testing. Moreover, 588 (74.1%) of the men provided urine samples for testing other infections. The prevalence of HSV-2 and HIV-1 among the study participants was 39.2% and 6.5%, respectively. The rate of HIV-1 in HSV-2 positive men was 11%, which is 3.2 times higher than the rate of HIV-1 in HSV-2 negative men. The rate of HSV-2 in HIV-1 positive men was 68%, 1.8 times higher than the rate of HSV-2 in HIV-1 negative men. The Pearson Chi-square test of interdependence between HSV-2 and HIV-1 was highly significant thereby indicating that men who were infected by one of the infections were likely to be infected by the other as well (`phi`-coefficient $= 0.15$, p-value < 0.0001). The median age of the study participants was 36 years (IQR $=$ 30–43). About two-thirds of the participants had only pre-secondary education, and one-third had at least secondary education. Thirty-three percent of them were classified as alcohol abusers. The median age at sexual onset was 18 years (IQR $=$ 16–20). The reported number of sexual partners in the last three years ranged from 1 to 10, with the majority of the men (78%) reported having one sex partner. Eighty percent of the men never used condom during sexual intercourse in the prior 12 months. Eleven percent of the men tested positive for at least one STI (other than HIV-1 and HSV-2) and eight percent reported STIs symptoms.

Preliminary analysis was done to identify the potential risk factors of HSV-2 and HIV-1 infections. All the variables that were found to be significantly associated with HIV-1 and HSV-2 at 0.10 levels in the preliminary analysis were further considered for multivariate analysis. Number of sexual partners (part) was associated with both infections. Age (age) and alcohol abuse (alco) were associated with HSV-2. Symptoms of STIs (symp) were associated with HIV-1. We did a multivariate analysis using these variables as covariates in the marginal (1–2) and joint response (3) models. The marginal models of HSV-2 and HIV-1 are given as $\text{logit}\{\text{E}(Y_1)\} = (\beta_{10} + \beta_{11}\text{age} + \beta_{12}\text{part} + \beta_{13}\text{alco})$ and $\text{logit}\{\text{E}(Y_2)\} = (\beta_{20} + \beta_{21}\text{part} + \beta_{22}\text{symp})$ where Y_1 and Y_2 denote HSV-2 and HIV-1 infection status of the jth subject as defined in Section 2. Similarly, using Equation (3), the joint response model is given as $\text{logit}\{\text{E}(Y_{ij})\} = d_{1i}(\tilde{\beta}_{10} +$

Table 4. Risk factors for HSV-2 and HIV-1 among men in Moshi, Northern Tanzania.

Parameter	OR	SE	95% CI	SE Ratio
Joint Model				
HSV-2				
Age	1.0583	0.0137	1.0314–1.0853	0.9786
Multiple partners	1.2614	0.3794	0.5168–2.0061	0.9055
Alcohol abuse	1.5965	0.3528	0.9040–2.2891	0.9958
HIV-1				
Multiple partners	2.1117	0.8427	0.4574–3.7660	0.8725
STIs Symptoms	3.3654	1.3743	0.6677–6.0631	0.9646
Marginal Models				
HSV-2				
Age	1.0617	0.0140	1.0342–1.0893	
Multiple partners	1.3560	0.4190	0.5324–2.1797	
Alcohol abuse	1.3718	0.3543	0.6774–2.0662	
HIV-1				
Multiple partners	2.3418	0.9658	0.4448–4.2388	
STIs Symptoms	3.4218	1.4247	0.6235–6.2200	

$\tilde{\beta}_{11}$age + $\tilde{\beta}_{12}$part + $\tilde{\beta}_{13}$alco + b_j) + d_{2i}($\tilde{\beta}_{20}$ + $\tilde{\beta}_{21}$part + $\tilde{\beta}_{23}$symp + b_j) where $d_{1i} = 1$ for HSV-2 and $d_{2i} = 1$ for HIV-1.

The results of our multivariate analysis are reported in Table 4. Odds ratios together with their corresponding standard errors and Wald-type 95% confidence intervals are presented. Moreover, the ratios of standard errors of the joint response model to the marginal model are presented in Table 4. Despite the large sample size ($n = 799$) and relatively lower correlation between the two outcomes ($\rho = 0.15$), there were efficiency gains in the joint model compared to the marginal models. The standard errors of the parameter estimates in the joint response model were smaller compared to the corresponding standard errors from the marginal models. However, both models lead to the same conclusion. The results of the joint response in Table 4 indicate that older age (41+) was associated with higher prevalence of HSV-2. The odds of having HSV-2 in older men were 1.06 times higher than that of younger men (OR = 1.06; 95% CI: 1.03–1.09). Men with two or more partners were associated with higher rates of HSV-2 (OR = 1.26; 95% CI: 0.52–2.01) and HIV-1 (OR = 2.11; 95% CI: 0.46–3.77) infections. Alcohol abuse was associated with higher prevalence of HSV-2 (OR = 1.60; 95% CI: 0.90–2.29). Men who had reported STIs symptoms were more likely to have HIV-1. The odds of having HIV-1 in men with STIs symptoms were 3.37 times higher than that of men without STIs symptoms (OR = 3.37; 95% CI: 0.67–6.06).

5. Conclusion

In this paper, we developed a joint response model for HIV-1 and HSV-2 infections. Our study aimed to assess the determinants of HSV-2 and HIV-1 infections among men in sub-Saharan Africa. Various socio-demographic and sexual characteristics as well as biological risk factors of HSV-2 and HIV-1 were considered. Seven hundred and ninety-four men enrolled in a community-based survey from the Moshi urban district of northern Tanzania were included in the study. Tanzania, with a population of 44.9 million, has about 5.1% of its adults infected with HIV. HIV prevalence has fallen in Tanzania over the past decade, it declined from 7% in 2004 to 5.1% in 2012 [31]. The prevalence rates of HSV-2 and HIV-1 among the study participants were 39.2% and 6.5%, respectively. The rate of HIV-1 in HSV-2 positive men was 3.2 times higher than the rate of HIV-1 in HSV-2 negative men. Moreover, the rate of HSV-2 in HIV-1 positive men was 1.8 times higher than the rate of HSV-2 among HIV-1 negative men. These indicate that men who were infected by one of the infections were more likely to be infected by the other as well. Thus, we used a joint response model that accommodates the interdependence between the two infections in assessing the potential risk factors for HSV-2 and HIV-1 infections. Classical models that ignore the dependence between the two infections result in biased estimates. We performed simulation studies to examine the performance of the joint response model compared to the separate analyses of the two outcomes. Our simulation studies suggest that joint modeling of the two outcomes provides better gains in efficiency in the parameter estimates. The gains in efficiency were more noticeable for smaller sample sizes, higher correlations between the two outcomes and larger proportions of missing data in the study outcomes.

We found high prevalence of HSV-2 among older men. The odds of HSV-2 in older men were 1.06 times higher than in younger men (OR = 1.06; 95% CI: 1.03–1.09). This is of special interest with regard to the control of the HIV epidemic in sub-Saharan Africa where sexual mixing of younger women with older men has been recognized to pose a high STIs risk for women and girls [13,16]. Maintaining sexual relationships with older men is common in the region primarily for economic benefits. Alcohol abuse and multiple sexual partners were associated with higher prevalence of HSV-2 and HIV-1. These findings are consistent with previous studies in the sub-region [4,5,15]. The reported symptoms of STIs were associated with higher prevalence of HIV-1

in our cohort, indicating that the syndromic approach can be used as a proxy for the laboratory test in a resource-poor setting where diagnostic laboratory tests are not always available.

In sub-Saharan Africa, where HIV-1 prevalence is the highest, the prevalence of HSV-2 is substantially high. The high prevalence of HSV-2 and the association between HSV-2 infection and HIV-1 acquisition could play a major role in the spread of HIV-1 in the region. Screening for HSV-2, HSV-2 treatment and addressing the risk factors of HSV-2 could serve as preventive measures for HIV-1 in sub-Saharan Africa where HIV-1 treatment is not readily available. Moreover, in the region, gender inequality is a central feature and sexual decision-making is the prerogative of men. Sexually active men are an important bridge population with regard to the HIV epidemic in the region. Effective control and prevention programs for HIV-1 in the region should have components aimed at addressing modifiable risk behaviors in men.

The present study has several strengths compared to previous studies. First, it is a large population-based study, enhancing its generalizability. Most studies on the risk factors of HSV-2 or/and HIV-1 have been carried out in particular groups that are not representative of the overall population, such as STD or antenatal clinic attendees, or men who have sex with men. Second, the correlation between the two infections was taken into consideration in our analytic approach thereby yielding unbiased estimates. Previous studies of HIV and HSV-2 co-infection modeled each infection separately ignoring the potential biological association between the two infections. There is a loss of efficiency in models which ignore the extra information contained in the correlation between the infections. However, our study has some inherent limitations that should be considered when interpreting the results. A strong case for causality cannot be made given the cross-sectional data, but findings can add to knowledge in the field and indicate where additional research could be directed. Longitudinal studies are needed to elucidate the nature of the relationship between HSV-2/HIV-1 infections and their risk factors. Moreover, the accuracy of the information given may be affected by the sensitive nature and potentially embarrassing information regarding high-risk behaviors (age at first sex, number of partners, alcohol abuse and condom use). Thus, the self-reported high-risk behaviors might have introduced social desirability and recall biases and were likely to be underreported.

Acknowledgments

The author would like to thank the editor and two referees for their constructive comments and suggestions. The author also thanks the women of Moshi Urban District of Tanzania for their participation and staff at the National Bureau of Statistics in Tanzania for conducting the survey. The author is grateful for the financial support from the Harvard University Center for AIDS Research (HU CFAR NIH/NAIDS P30-AI 060354).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] L.J. Abu-Raddad, A.S. Magaret, C. Celum, A. Wald, I.M. Longini, S.G. Self, and L. Corey, *Genital herpes has played a more important role than any other sexually transmitted infection in driving HIV prevalence in Africa*, Plos one 3(5) (2008), p. e2230.

[2] S. Bianconcini, *Asymptotic properties of adaptive maximum likelihhod estimators in latent variable models*, Bernoulli 3 (2014), pp. 1507–1531.

[3] S. Cagnone and P. Monari, *Latent variable models for ordinal data by using the adaptive quadrature approximation*, Comput. Stat. 28(2) (2013), pp. 597–619.

[4] L. Chen, P. Jha, B. Stirling, S.K. Sgaier, T. Daid, R. Kaul, and N. Nagelkerke, *Sexual risk factors for HIV infection in early and advanced HIV epidemics in sub-Saharan Africa: Systematic overview of 68 epidemiological studies*, PLoS One 2(10) (2007), p. e1001.

[5] R.L. Cook and D.B. Clark, *Is there an association between alcohol consumption and sexually transmitted diseases? A systematic review*, J. Sexually Transmitted Diseases 32(3) (2005), pp. 156–164.

[6] L. Corey, A. Wald, C.L. Celum, and T.C. Quinn, *The effects of herpes simplex virus-2 on HIV-1 acquisition and transmission: a review of two overlapping epidemics*, J. Acquired Immune Deficiency Syndromes 35(5) (2004), pp. 435–445.

[7] A.L. Cunningham, R.R. Turner, A.C. Miller, M.F. Para, and T.C. Merigan, *Evolution of recurrent herpes simplex lesions. An immunohistologic study*, J. Clin. Invest. 75(1) (1985), pp. 226–233.

[8] G. Fitzmaurice, N. Laird, and J. Ware, *Applied Longitudinal Analysis*, Wiley, Hoboken, NJ, 2004.

[9] D.T. Fleming and J.N. Wasserheit, *From epidemiological synergy to public health policy and practice: The contribution of other sexually transmitted diseases to sexual transmission of HIV infection*, Sexually Transmitted Infections 75(3) (1999), pp. 3–17.

[10] E.E. Freeman, K.K. Orroth, R.G. White, J.R. Glynn, R. Bakker, M.-C. Boily, D. Habbema, A. Buvé, and R. Hayes, *Proportion of new HIV infections attributable to herpes simplex 2 increases over time: Simulations of the changing role of sexually transmitted infections in sub-Saharan African HIV epidemics*, Sexually Transmitted Infections 83 (2007), pp. i17–i24.

[11] S.R. Galvin and M.S. Cohen, *The role of sexually transmitted diseases in HIV transmission*, Nat. Rev. Microbiol. 2(1) (2004), pp. 33–42.

[12] M. Ghebremichael, D. Hadera, and E. Paintsil, *Deciphering the epidemic synergy of herpes simplex virus type 2 (HSV-2) on human immunodeficiency virus type 1 (HIV-1) infection among women*, BMC Res Notes 5(451) (2012), pp. 1–7.

[13] S. Gregson, C.A. Nyamukapa, G.P. Garnett, P.R. Mason, T. Zhuwau, M. Caraël, S.K. Chandiwana, and R.M. Anderson, *Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural Zimbabwe*, Lancet 359(9321) (2002), pp. 1896–1903.

[14] R.V. Gueorguieva and G. Sanacora, *Joint analysis of repeatedly observed continuous and ordinal measures of disease severity*, Stat. Med. 25 (2006), pp. 1307–1322.

[15] S.C. Kalichman, L.C. Simbayi, M. Kaufman, D. Cain, and S. Jooste, *Alcohol use and sexual risks for HIV/AIDS in sub-Saharan Africa: Systematic review of empirical findings*, Prev. Sci. 8(2) (2007), pp. 141–151.

[16] R.J. Kelly, R.H. Gray, N.K. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, and M.J. Wawer, *Age differences in sexual partners and risk of HIV-1 infection in rural Uganda*, J. Acquired Immune Deficiency Syndromes 32(4) (2003), pp. 446–451.

[17] D.M. Koelle, H. Abbo, A. Peck, K. Ziegweid, and L. Corey, *Direct recovery of herpes simplex virus (HSV)-specific T lymphocyte clones from recurrent genital HSV-2 lesions*, J. Infectious Diseases 169(5) (1994), pp. 956–961.

[18] U. Larsen, J. Mlay, S. Aboud, R. Ballard, N.E. Sam, J.F. Shao, and S.H. Kapiga, *Design of a community-based study of sexually transmitted infections/HIV and infertility in an urban area of northern Tanzania*, Sexually Transmitted Diseases 34(1) (2007), pp. 20–24.

[19] S.R. Lipsitz, N.M. Laird, and D.P. Harrington, *Maximum likelihood regression methods for paired binary data*, Stat. Med. 9 (1990), pp. 1517–1525.

[20] Q. Liu and D. Pierce, *A note on Gauss–Hermite quadrature*, Biometrika 81(3) (1994), pp. 624–629.

[21] C. Ludema, S.R. Cole, C. Poole, H. Chu, and J.J. Eron, *Meta-analysis of randomized trials on the association of prophylactic acyclovir and HIV-1 viral load in individuals coinfected with herpes simplex virus-2*, AIDS 25 (2011), pp. 1265–1269.

[22] P. Mayaud and D. McCormick, *Interventions against sexually transmitted infections (STI) to prevent HIV infection*, Br. Med. Bull. 58 (2001), pp. 129–153.

[23] D. Mayfield, G. McLeod, and P. Hall, *The CAGE questionnaire: validation of a new alcoholism screening instrument*, Am. J. Psychiatry 131(10) (1974), pp. 1121–1123.

[24] C. McCulloch, *Joint modelling of mixed outcome types using latent variables*, Stat. Methods Med. Res. 17(1) (2008), pp. 53–73.

[25] A. Ouedraogo, N. Nagot, L. Vergne, I. Konate, H.A. Weiss, M.-C. Defer, V. Foulongne, A. Sanon, J.-B. Andonaba, M. Segondy, P. Mayaud, and P. Perre, *Impact of suppressive herpes therapy on genital HIV-1 RNA among women taking antiretroviral therapy: A randomized controlled trial*, AIDS 20(18) (2006), pp. 2305–2313.

[26] S. Rabe-Hesketh, A. Skrondal, and A. Pickles, *Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects*, J. Econ. 128 (2005), pp. 301–323.

[27] T. Schacker, J. Zeh, H.L. Hu, E. Hill, and L. Corey, *Frequency of symptomatic and asymptomatic herpes simplex virus type 2 reactivations among human immunodeficiency virus-infected men*, J. Infectious Diseases 178(6) (1998), pp. 1616–1622.

[28] T. Schacker, J. Zeh, H. Hu, M. Shaughnessy, and L. Corey, *Changes in plasma human immunodeficiency virus type 1 RNA associated with herpes simplex virus reactivation and suppression*, J. Infectious Diseases 186(12) (2002), pp. 1718–1725.

[29] A. Skrondal and S. Rabe-Hesketh, *Generalized Latent Variable Modeling*, Chapman & Hall, New York, 2004.

[30] A. Teixeira-Pinto and S.-L.T. Normand, *Correlated bivariate continuous and binary outcomes: Issues and applications*, Stat. Med. 28 (2009), pp. 1753–1773.
[31] The United Republic of Tanzania. Tanzania in Figures 2012. National Bureau of Statistics, Ministry of Finance June, 2013.
[32] UNAIDS, Report on the global AIDS epidemic. Joint United Nations Programme on HIV/AIDS (UNAIDS), 2013.
[33] S.L. Zeger, K.-Y. Liang, and P.S. Albert, *Models for longitudinal data: A generalized estimating equation approach*, Biometrics 44 (1988), pp. 1049–1060.