Published in final edited form as:

J Am Stat Assoc. 2015 June 1; 110(510): 472–485. doi:10.1080/01621459.2014.991394.

Analysis of longitudinal multivariate outcome data from couples cohort studies: application to HPV transmission dynamics

Xiangrong Kong,

Department of Epidemiology and Department of Biostatistics

Mei-Cheng Wang, and

Department of Biostatistics

Ronald Gray

Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University

Abstract

We consider a specific situation of correlated data where multiple outcomes are repeatedly measured on each member of a couple. Such multivariate longitudinal data from couples may exhibit multi-faceted correlations which can be further complicated if there are polygamous partnerships. An example is data from cohort studies on human papillomavirus (HPV) transmission dynamics in heterosexual couples. HPV is a common sexually transmitted disease with 14 known oncogenic types causing anogenital cancers. The binary outcomes on the multiple types measured in couples over time may introduce inter-type, intra-couple, and temporal correlations. Simple analysis using generalized estimating equations or random effects models lacks interpretability and cannot fully utilize the available information. We developed a hybrid modeling strategy using Markov transition models together with pairwise composite likelihood for analyzing such data. The method can be used to identify risk factors associated with HPV transmission and persistence, estimate difference in risks between male-to-female and female-to-male HPV transmission, compare type-specific transmission risks within couples, and characterize the inter-type and intra-couple associations. Applying the method to HPV couple data collected in a Ugandan male circumcision (MC) trial, we assessed the effect of MC and the role of gender on risks of HPV transmission and persistence.

Keywords

Alternating logistic regression; Pairwise likelihood; Clustered binary data; Composite likelihood; Markov transition model

1 INTRODUCTION

Correlated data are common in clustered studies where cross-sectional correlation exists among outcomes within a cluster, or in longitudinal studies where temporal correlation exists among the repeated measurements. In more complicated situations, correlation may arise from different sources. For example, Geys et al. (1999) considered outcomes of external, visceral, and skeletal malformations in fetuses born to mice dosed with an industrial chemical where directionless correlation existed both between the multiple

malformation outcomes within a fetus and between the fetuses from the same litter. Liang and Zeger (1989) and O'Brien and Fitzmaurice (2004) considered a univariate binary outcome repeatedly reported on family members (parents and children) where the outcome of the different family members was interrelated, together with temporal correlation due to the longitudinal follow-up. Here we consider a specific situation where multivariate binary outcomes are repeatedly measured on each member of a cluster, leading to the simultaneous presence of temporal correlation and cross-sectional inter-outcome and intra-cluster correlations. The motivating context is research on genital human papillomavirus (HPV) transmission dynamics using couple cohort studies.

HPV infection is a common sexually transmitted infection (STI), with high prevalence in both developing and developed countries (Morse et al., 2003). More than 40 types have been identified as causing genital infections (Morse et al., 2003), and these different HPV types have different etiological characteristics and exhibit different clinical symptoms. In particular, fourteen oncogenic high risk types (HR-HPV) (HPV-16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) are known to be on the causal pathway to anogenital cancers, including penile cancer in men and cervical cancer in women.

Epidemiological studies have largely focused on gender specific HPV infections, and a few statistical methods have been proposed to address various analytical challenges from such studies. For example, to deal with possibly misclassified outcomes due to imperfect diagnostic technologies, Bureau et al. (2003) applied two-state hidden Markov models (HMM) to estimate intensities of acquisition and clearance of HPV 16 and 18 in women enrolled in an HPV natural history study. The use of HMM also allows estimation of misclassification rates in the observed testing results. Kang and Lagakos (2004) discussed issues of using HPV infection as a surrogate endpoint in HPV vaccine trials and suggested a multi-state model for assessing vaccine effects on the transitions between HPV infection states (positive or negative) and the clinical end point of cervical intraepithelial neoplasia (CIN). Kang and Lagakos (2007) further used a continuous time multi-state semi-Markov process to model the data from an HPV vaccine trial in women, and transition intensity functions between states of HPV-16 infection and the clinical state of CIN were used to parametrize the process. Possible misclassification rates were also considered but as pre-specified parameters in the likelihood function. More recently, Mitchell et al. (2011) proposed a discrete-time semi-Markov model that avoids the parametric assumptions in Kang and Lagakos (2007), and provides a modeling framework that can accommodate different operational definitions of HPV persistence used by different HPV researchers. All these methods apply to data from an individual HPV type. Previous work addressing concurrent infections with multiple HPV types include parametric frailty models with arbitrary censoring for the outcome of HPV clearance (Kong et al., 2010a) and a semi-parametric approach combining a Markov transition model with the GEE technique for the purpose of identifying risk factors for HPV acquisition in men (Kong et al., 2010b).

While gender-specific infections have been well studied, recent epidemiological studies have focused on elucidating transmission dynamics in heterosexual couples (WHO International Agency for Research on Cancer, 2007). Knowledge about HPV transmission between sex partners is important for understanding HPV epidemiology and estimating the

population effectiveness of HPV vaccines (Trottier and Franco, 2006; Burchell et al., 2010b,a; Hernandez et al., 2008; Franco et al., 2006; WHO International Agency for Research on Cancer, 2007). For example, couple studies provide a unique opportunity to estimate levels of and differences between male-to-female and female-to-male HPV transmission risks. Such information is critical for cost effectiveness research on HPV prevention strategies, and will aid policy decisions about the need for vaccination of boys and young men (Palefsky, 2010; Kim and Goldie, 2009).

Technological advances in DNA testing have allowed the simultaneous detection and genotyping of multiple HPV infections from a single DNA amplification assay. As a result, longitudinal studies of HPV in couples typically include multiple infection outcomes from each partner, measured at each study visit. Considering the underlying transmission dynamics, the correlation structure characterizing the observed outcomes from such studies is potentially very complex. Analyses that fail to account for this complexity are likely to lead to incomplete or invalid scientific conclusions. One published HPV couple transmission study followed 25 monogamous heterosexual Hawaiian couples for an average of 7.5 months, with sampling at 2-month intervals from multiple anatomic sites (Hernandez et al., 2008). The study reported transmission between partners and transmission between different sampling sites within individuals (i.e. auto-inoculation). It was concluded that HPV transmission “may” differ by type and the rates of female-to-male transmission “were substantially higher than” those of male-to-female transmission. However, these conclusions were based on a summary of the numbers of transmission events and on using the Poisson distribution to estimate confidence intervals for transmission rates. The fact that multiple transmission events often occurred from the same index partner was ignored, resulting in incorrect estimates of the standard errors. Information regarding the likelihood of observing multiple events on the same individual or within the sample couple was also discarded. Moreover, there was no formal statistical comparison between female-to-male and male-to-female transmission risks or between transmission rates for different types, leaving unknown the magnitude and significance of the differences between gender- or type-specific transmission risks.

In this article, we present an analytical framework for longitudinal multivariate outcome data collected on clusters and apply it to analyze HPV data from couple cohort studies. In what follows we first briefly review basic statistical methods for correlated data analysis. Section 3 then describes the typical data structure from HPV couple cohort studies and the analytical challenges it poses, and introduces the HPV couple dataset collected from the Rakai male circumcision trial in Uganda and the relevant scientific questions of interest. Section 4 presents our proposed analytical method and discusses parameter interpretations and parameterization for the pairwise composite likelihood. Estimation and computation are described in Section 5. This method is applied to the Rakai MC trial data and results of the applications are shown in Section 6. Section 7 discusses potential uses of the proposed analytical tool and its limitations.

2 A BRIEF REVIEW OF METHODS FOR CORRELATED BINARY DATA

Statistical methods for analyzing correlated binary data have been extensively studied, and can be summarized into a few major classes, including marginal models based on generalized estimating equations (GEE) (Liang and Zeger, 1986), random effects models (Breslow and Clayton, 1993; Ten Have and Morabia, 1999; Chib and Jeliazkov, 2006), and conditional models (including transitional i.e. Markov or regressive model as a special case) (Bonney, 1987; FitzGerald and Knuiman, 1998). A family of hybridized marginal-conditional models has also been proposed in the literature and summarized in Molenberghs and Verbeke (2005). This approach specifies both a regression model for the first and second-order marginal means and a model for higher order conditional odds ratios (COR) (i.e. conditioning on other outcome variables), and parameter estimation uses the score equation from the likelihood function parametrized by the first and second order marginal means and higher order CORs. As a special case, Carey et al. (1993)'s alternating logistic regression (ALR) specifies a model for the first order marginal mean and a logistic regression model for the second order marginal odds ratios where an outcome is modeled conditioning on another outcome.

For the analysis of multivariate data, the framework of composite likelihood (CL), a terminology introduced by Lindsay (1988), has recently gained more attention (Varin, 2008; Varin et al., 2011). The CL is essentially a pseudo-likelihood object composed by taking the product over low dimension marginal or conditional densities. Similar to maximum likelihood estimation (MLE), inference on the unknown parameters is carried out by solving the CL score equation (i.e. maximizing the CL). Under certain regularity conditions, consistency and asymptotic normality hold with the usual Fisher information replaced by the Godambe information (also called sandwich information) (Lindsay, 1988; Varin, 2008). The CL method provides an objective function for parameter estimation based only on low dimensional densities, and is particularly attractive in multivariate problems where the analytical form of the full joint likelihood function is not available or computationally infeasible (Varin, 2008). When using the CL framework, pairwise likelihood is often considered (Heagerty and Lele, 1998; Le Cessie and Van Houwelingen, 1994; Hjort and Varin, 2007; Zhao and Joe, 2005; Parner, 2001; Kuk and Nott, 2000) where the product of all possible pairwise marginal or conditional likelihood functions comprises the CL. For example, Geys et al. (1999)'s analysis of the multiple malformation outcomes in mice utilized the pseudolikelihood composed by bivariate conditional densities.

While pairwise likelihood can be used for estimation of both the first-order marginal and bivariate correlation (association) parameters, Kuk (2007) proposed a modified pairwise CL method whereby a combination of estimating equations based on both univariate and bivariate score functions is used to draw inference on the marginal as well as the correlation parameters. The resulting estimates for the marginal parameters are more efficient than those using the pairwise likelihood alone and are robust to misspecification of the pairwise likelihood function. A similar idea was also suggested by Firth (1992).

In practice, the best model can only be determined with a clear understanding of the scientific question being addressed. With a high dimensional response vector that may

simultaneously contain inter-type, intra-couple, and temporal correlations from the motivating HPV couple data, existing approaches for correlated data analysis cannot be directly applied. Here we present a statistical modeling framework for rigorous scientific inferences from HPV couple cohort studies. We propose a semi-parametric hybrid model for couple-level HPV data analysis that combines a Markov transition model with Kuk (2007)'s modified pairwise composite likelihood approach to deal with the different kinds of correlations.

Use of Markov transition models in longitudinal data analysis has been limited, with one of the reasons being that the effect estimate may be smaller after conditioning on the past history of the response (Fitzmaurice et al., 2009, Chapter 1). With this application, however, we hope to show that for certain epidemiology studies on recurrent diseases, such as HPV infection or herpes simplex virus type 2 (HSV-2) reactivation (Cherpes et al., 2005), such issue is no longer a disadvantage as the effects of the covariates after conditioning on the past are what is of scientific interest. Transition models in fact provide a natural modeling tool as they directly reflect the scientific outcome event of interest, namely incidence, clearance, reactivation, or transmission which is defined by conditioning on past responses.

3 TYPICAL DATA STRUCTURE FROM HPV COUPLE COHORT STUDIES

3.1 The multifaceted correlation structure and analytical challenges

In HPV couple cohort studies, couples are followed over time with pre-determined concurrent follow-up visits (e.g., every six months). At each visit, biological specimens are collected from male and female partners respectively. Each specimen is then tested using an HPV-DNA detection assay. Current technology allows the simultaneous detection and genotyping of multiple HPV infections from a single assay, such as the Roche HPV Linear Array (Roche Diagnostics, Indianapolis, IN). Human β -globin gene amplification is included as an internal control for sample adequacy. This PCR-based assay has high sensitivity and specificity and can detect 37 common HPV types (Halfon et al., 2010), including 23 low risk types (HPV-6, 11, 26, 40, 42, 53, 54, 55, 61, 62, 64, 67, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39, and CP6108), and the aforementioned 14 high risk oncogenic types. The HPV outcome data can hence be described using 37 Bernoulli variables.

The complicated correlation structure in the HPV couple testing data is determined by the biology of HPV and can be characterized into two distinct types. Longitudinal data suggest that there may be temporal correlation between visits where the current testing result of an individual may depend on both his/her own and the sex partner's infection history. In particular, HPV infection is commonly transient with an estimated median duration of less than 6 months in men for prevalent infections (i.e. infections detected at study baseline) (Giuliano et al., 2008; Hernandez et al., 2008) and 8 months in women for incident infections (i.e. new infections detected during study follow-up) (Morse et al., 2003). Thus it can be observed in a longitudinal study that recurrent events of new detection or persistence of the same HPV type happen in the same individual.

There may be also directionless correlations in the data collected at each visit, which can be further differentiated into two sources: within-individual (inter-type) and between-

individuals (intra-couple). Sexual contact is the major transmission route of genital HPV. Within an individual, detection of one type may increase the odds of detection of another type, thus within-individual testing results on the multiple types may be highly correlated. In addition, infection status of an individual strongly influences his/her partner's status (Burchell et al., 2010a), and thus the between-individual HPV results may be also highly correlated. Finally, in data collected from polygamous relationships, women's sexual relationships with a common husband might also introduce correlation between infection statuses of the multiple wives.

These different correlations do not operate in isolation, and this interrelated multifaceted correlation structure presents analytical challenges. Nevertheless, it is also integral to the scientific assessment of HPV transmission dynamics: a transmission event within a couple is often operationally defined, based on the observed testing results from DNA assays, as the detection of a new type at time t in an individual who was previously negative for this type and whose partner was positive for this type at time $t - 1$ (Burchell et al., 2010b; Hernandez et al., 2008); and persistence may be operationally defined as the detection of the same HPV type in two (or more) consecutive visits. The analytical handling of the correlations has to allow interpretable scientific inference on HPV transmission.

3.2 The HPV couple cohort from the Rakai male circumcision trial

The Rakai male circumcision (MC) trial was conducted during 2003–07 in Rakai District of Uganda, and the primary outcome of interest was to evaluate the efficacy of MC on preventing HIV acquisition in men, with a secondary outcome to study the effect of MC on HPV infection in men and their female partners. The trial enrolled 4996 HIV-negative uncircumcised men (Gray et al., 2007) who were randomized into immediate circumcision (intervention arm) or circumcision delayed for 24 months (control arm). They were followed at 6, 12 and 24 months to determine HIV and STI acquisition, and provided penile swabs at enrollment and at each follow-up visit for detection of HPV. Consenting female partners of men who were married or in long-term consensual relationships were concurrently enrolled in a parallel study and followed up at 12 and 24 months (note a man in polygamous relationship may have multiple female partners enrolled). At each visit, women provided blood samples for HIV and STI testing, and self-collected vaginal swabs for HPV detection. Roche HPV Linear Array (Roche Diagnostics, Indianapolis, IN) was used for HPV detection and genotyping. Additionally, at each visit, interview data on sociodemographic characteristics, sexual behaviors, and health status were collected from both men and women. Therefore, couple level data concurrently measured at baseline, 12, and 24-months from the Rakai MC trial were available to examine the HPV transmission dynamics.

Due to the original MC trial design, the Rakai couple HPV data are subject to the limitation of a long sampling interval (1-year) where transmission events that occurred during the year may have already cleared by the next visit and thus are unobserved through the data. Moreover, as in other studies, the participants (especially men) may have other sex partners that were not captured in the study, and thus there may be transmission events from unidentified sources. Nevertheless, embedded in the design of a randomized controlled trial, these couple-level longitudinal HPV data provide a unique opportunity to estimate the

treatment effects of MC on preventing male-to-female and female-to-male HPV transmission and on persistence of a pre-existing HPV infection. We will also compare the male-to-female and female-to-male HPV transmission risks and the risk of persistence between genders, broadening our knowledge of HPV epidemiology in a typical rural sub-Saharan African setting with no access to HPV vaccines. In addition, it is of secondary interest to characterize the correlations (or associations) between partners and between HPV types within an individual. With the sexually transmitted nature of genital HPV infection, we expect strong associations in the HPV status between partners and within an individual (i.e. between types).

There were 640 couples enrolled in the Rakai MC trial and assayed for HPV testing, including 336 men and their 366 female partners from the intervention arm and 304 men and their 323 female partners from the control arm. The range of number of female partners per household was 1–3, with the majority of households being monogamous (Note: we use “household” to denote the unit composed by a man and his female partners. In a polygamous relationship, however, the female partners do not necessarily live in the same abode in Rakai). The intervention and control arms were comparable in sociodemographic and sexual risk behaviors for both the men and women. Figure 1 presents the prevalence of oncogenic HPV infection in men and women by randomization arm at each visit. Non-trivial proportions of men and women were infected with multiple oncogenic HPV types. At enrollment, HPV prevalence was similar between men and women. A decreasing trend in HPV prevalence was observed in both men and women in the intervention arm, with a larger decline in men. Prevalence also decreased in the control arm men (i.e. uncircumcised), possibly because of reduced risk behaviors with the intensive health education and counseling men received during the trial (Gray et al., 2007); but the prevalence in women in the control arm did not change.

There were 338 discordant infections at enrollment (i.e. infections carried by only one partner and thus able to be transmitted to the other partner) observed in 211 couples and 186 discordant infections in 122 couples observed at year 1. Table 1 summarizes the number of infections that were transmitted to the other partner by the next visit, stratifying by gender and MC status. Also presented are the naive odds ratios (ORs) comparing transmission odds between couples with circumcised men and couples with uncircumcised men. Note that while such ORs provide a simple summary of transmission risk by circumcision status, they ignore the possible temporal, inter-type and intra-couple correlations, resulting in potentially invalid confidence intervals.

4 A HYBRID MODELING APPROACH FOR STUDYING HPV TRANSMISSION DYNAMICS WITHIN COUPLES

4.1 Notation

Let $Y_{ijg}(t)$ denote the binary outcome observed at the t th visit on HPV type g of partner j from household i , where $i = 1, 2, \dots, n$ indexes household, with n being the number of households in the study; $j = 1, 2, \dots, J_i$ indexes the partner, with J_i being the number of sex partners in household i ; $g = 1, 2, \dots, G$ indexes type with G being the number of types; and $t =$

0, 1, ..., K, with K being the number of follow-up visits. K is normally pre-specified by study design and fixed for all households. Different households are assumed to be independent from each other.

4.2 Model formulation-Monogamous household

For ease of illustration, we first consider monogamous couples. Thus, a household has only one male and one female partner, and $J_i = 2$ for all i . For couple i , the observed outcomes are $\mathbf{y}_i(t) = (\mathbf{y}_{i1}^T(t), \mathbf{y}_{i2}^T(t))^T$ (bold font for vector) where $\mathbf{y}_{ij}(t) = (y_{ij1}(t), y_{ij2}(t), \dots, y_{ijG}(t))^T$ is the $G \times 1$ vector for the outcomes of all types for partner j at visit t . The joint probability function for household i is:

$$P(\mathbf{Y}_{i1}(0)=\mathbf{y}_{i1}(0), \dots, \mathbf{Y}_{i1}(K)=\mathbf{y}_{i1}(K); \mathbf{Y}_{i2}(0)=\mathbf{y}_{i2}(0), \dots, \mathbf{Y}_{i2}(K)=\mathbf{y}_{i2}(K)) \quad (1)$$

It is important to note that the operational definition used by the HPV field for “transmission” (section 3.1) is in fact based on a conditioning argument, hence the conditional probability (i.e. transitional probability) provides a natural mathematical modeling tool and can be used to model the temporal correlation. Using the basic conditional probability property, the joint probability function in (1) can be factorized as:

$$P(\mathbf{y}_{i1}(0), \mathbf{y}_{i2}(0)) \cdot \prod_{t=1}^K P(\mathbf{y}_{i1}(t), \mathbf{y}_{i2}(t) | \mathbf{y}_{i1}(t-1), \mathbf{y}_{i2}(t-1); \dots; \mathbf{y}_{i1}(0), \mathbf{y}_{i2}(0)) \quad (2)$$

Note that $P(\mathbf{y}_{i1}(0), \mathbf{y}_{i2}(0))$ relates to baseline prevalence information and accordingly does not influence the estimation of parameters for transmission or persistence.

Since HPV infection is often transient with a median duration of 6–8 months and one can reacquire the same type of infection (Morse et al., 2003), depending on the length of the follow-up interval, the conditional probability in (2) may be simplified by assuming that a couple’s infection statuses at t depend on their infection history observed at the previous q visits where the choice of q should consider the length of the follow-up interval. Specifically, the operational definition of a transmission event suggests that the dependence of an individual’s infection status at t on his/her and the partner’s infections statuses at the immediate prior visit (i.e. $q = 1$) is of primary interest. If the follow-up interval is relatively short, however, using $q > 1$ allows flexibility in the model to accommodate possible higher order dependence or operational definition of HPV “transmission” or “persistence” based on more than one prior visit. In addition, studies have suggested relatively high per-coital act transmission probability of HPV (Burchell et al., 2006a), thus we assume that after conditioning on an individual’s own past infection history (up to order q) and the sex partner’s testing results at the immediate prior visit, the index individual’s current testing status is independent of the partner’s results from earlier visits. Following these assumptions, we consider modeling the transition probability $p_{ijg(t)|(t-q)}$ where

$$p_{ijg(t)|(t-q)} = P[Y_{ijg}(t)=1 | \mathcal{A}_{ig}(t-1); \mathcal{B}_{ijg}; \mathcal{C}_{ij}(-g)] \quad (3)$$

where $\mathcal{A}_{ij}(t-1) = (y_{ijg}(t-1), y_{i(3-j)g}(t-1))$ records the partner and the index individual's own infection statuses at visit $t-1$ for type g (note individual j 's partner can be indexed by $3-j$); $\mathcal{B}_{ijg} = (y_{ijg}(t-2), \dots, y_{ijg}(t-q))$ records the index individual's own infection history with type g before $t-1$ and up to $t-q$; and $\mathcal{C}_{ij-g} = (y_{ij(-g)}(t-1), \dots, y_{ij(-g)}(t-q))$ records the individual's own infection history with all other types (denoted by $(-g)$). We then consider the regression model:

$$G(p_{ijg(t)|(t-q)}) = \mu + \beta_1 y_{ijg}(t-1) + \beta_2 y_{i(3-j)g}(t-1) + \beta_{12} y_{ijg}(t-1) \cdot y_{i(3-j)g}(t-1) + \alpha^T \mathbf{x}_{ij}(t) + \alpha_{\beta_1}^T \mathbf{x}_{ij}(t) \cdot y_{ijg}(t-1) + \alpha_{\beta_2}^T \mathbf{x}_{ij}(t) \cdot y_{i(3-j)g}(t-1) + H_1(y_{ijg}(t-2), \dots, y_{ijg}(t-q); \gamma_1) + H_2(y_{ij(-g)}(t-1), \dots, y_{ij(-g)}(t-q); \gamma_2) \quad (4)$$

where $G(\cdot)$ is an appropriate link function, and $\mathbf{x}_{ij}(t)$ is a vector of predictor variables measured at the t th visit. $H_1(\cdot)$ with parameters γ_1 is a function of infection statuses of type g before visit $t-1$, and $H_2(\cdot)$ with parameters γ_2 is a function of infection history with other types. For example, let $H_1(\cdot) = \gamma_1 \cdot I(\sum_{l=2}^q y_{ijg}(t-l) > 0)$ ($I(\cdot)$ is the indicator function), then γ_1 may reflect partial or full immunity due to an earlier infection with this type. Also, $H_2(\cdot)$ can simply be zero, since studies have suggested low cross-type immunity (Trottier and Franco, 2006; Burchell et al., 2006b). A non-zero $H_2(\cdot)$ however can be used to statistically explore whether there is cross-type protection.

4.3 Model formulation-Polygamous household

Extension of model (4) to accommodate polygamous relationships is straightforward and determined by the operational definition of "transmission". It is possible that a woman acquired a new HPV at t , her husband did not carry this type at $t-1$ but another wife in the household did. Since in Rakai, it is relatively rare for women to have a non-marital sex relationship, it is highly likely that the index wife transmitted the HPV to the husband who subsequently transmitted it to the other wife, and therefore we consider such case as a transmission event. Thus model (4) is extended to be:

$$G(p_{ijg(t)|(t-q)}) = \mu + \beta_1 y_{ijg}(t-1) + \beta_2 I(\sum_{j' \neq j} y_{ij'g}(t-1) > 0) + \beta_{12} y_{ijg}(t-1) \cdot I(\sum_{j' \neq j} y_{ij'g}(t-1) > 0) + \alpha^T \mathbf{x}_{ij}(t) + \alpha_{\beta_1}^T \mathbf{x}_{ij}(t) \cdot y_{ijg}(t-1) + \alpha_{\beta_2}^T \mathbf{x}_{ij}(t) \cdot I(\sum_{j' \neq j} y_{ij'g}(t-1) > 0) + H_1(y_{ijg}(t-2), \dots, y_{ijg}(t-q); \gamma_1) + H_2(y_{ij(-g)}(t-1), \dots, y_{ij(-g)}(t-q); \gamma_2) \quad (5)$$

The term $I(\sum_{j' \neq j} y_{ij'g}(t-1) > 0)$ is a summary of type g 's infection status at $t-1$ for all partners of individual j in household i , which may include partners of direct sexual relationships (e.g. husband-wife) or indirect relationships (e.g. wife 1-wife 2). Note if a man was observed to acquire a new HPV type at t and more than one of his wives carried the type at $t-1$, the linear array assay data cannot identify which partner transmitted the virus to the man. Thus the potential multiple transmission events from the multiple wives will only be counted once. In addition, due to the limitation of the long one-year sampling interval in the Rakai MC data, even though a transmission may be observed between the wives, the

intermediate transmission event from the index wife to the husband may not be captured by the data as the HPV may have cleared by t in the husband.

4.4 Interpretation of model parameters

Corresponding to the operational definition used in the literature (section 3.1), the event $\{Y_{ijg}(t) = 1 | y_{ijg}(t-1) = 0, I(\sum_{j' \neq j} y_{ij'g}(t-1) > 0) = 1\}$ represents a transmission event in the household. Thus if using the logit link and $H_1(\cdot) = H_2(\cdot) = 0$, for the reference group with covariates $\mathbf{x}_i(t) = \mathbf{0}$, $\mu + \beta_2$ is the log odds of transmission of a specific HPV type within a household; $\mu + \beta_1 + \beta_2 + \beta_{12}$ is the log odds of persistence of a pre-existing infection when at least one partner also had the same infection at $t-1$; $\mu + \beta_1$ is the log odds of persistence of a pre-existing infection when no partner had the virus detected at $t-1$; and μ is the log odds of a new HPV detection with an unknown source which may be a sex partner not captured during data collection, or a known partner but who had undetected infection at $t-1$ or acquired the infection between $t-1$ and t . Based on a hypothesized mechanism for the natural history of HPV infection, the observed new detection with unidentified sources may also be re-activation of a previous infection that has become latent and undetectable using current HPV assays (Gravitt, 2011).

The coefficients in α and $\alpha\beta$ reflect the effects of predictor variables, such as age, gender, HPV type, or sexual behaviors, on the odds of HPV detection at t , conditioning on the individual and partner's past testing results. Specifically, since detection of a new infection (including transmission from the known partner) and persistence of a pre-existing infection are different biological processes, the interaction terms of $\alpha\beta_1^T \mathbf{x}_{ij}(t)y_{ijg}(t-1)$ and $\alpha\beta_2^T \mathbf{x}_{ij}(t) \cdot y_{i(3-j)g}(t-1)$ in model (4) allow different effects of predictor variables on the two processes. For example, if an exposure variable x increases by one unit, its corresponding $\alpha + \alpha\beta_2$ is the log odds ratio (OR) of HPV transmission from the partner; and $\alpha + \alpha\beta_1$ is the log OR of persistence of a pre-existing infection when no partner carried such infection at $t-1$.

4.5 Constructing the composite likelihood

The $p_{ijg(t)|(t-q)}$ in (3) is the univariate mean conditioning on the past for type g , partner j at time t . Such conditioning naturally takes care of the temporal correlation in the data and directly reflects the scientific interest. The $p_{ijg(t)|(t-q)}$ is also marginal in relation to the outcomes at t of other types within partner j and outcomes of all types for his/her sex partner. Therefore, (4) can be viewed as a marginal and transitional regression model (for simplicity we term it as marginal-transition model) with marginal-transition parameters

$\theta = (\mu, \beta_1, \beta_2, \beta_{12}, \alpha^T, \alpha\beta_1^T, \alpha\beta_2^T, \gamma_1^T, \gamma_2^T)^T$. Note that after conditioning on the past, the current outcomes within an individual and between partners are still likely highly correlated given the sexually transmitted nature of HPV. Indeed, these associations between types within an individual and between partners are also of scientific interest (WHO International Agency for Research on Cancer, 2007).

Use of the full likelihood for inference is difficult given the high dimension of the response vector on multiple HPV outcomes at the cluster level. For example, for a monogamous couple tested at three visits, a multivariate binary vector with length 84 is needed to describe their oncogenic HPV outcomes. As an alternative, we consider using the CL method for

parameter estimation. Univariate CL (i.e. product of univariate densities conditioning on the past) is simple but does not allow estimation of the within-individual and between-partners correlations, whereas pairwise CL is informative for both the marginal-transitional and the association parameters. Nevertheless, we adopted the modified pairwise likelihood approach proposed by Kuk (2007) to estimate the two sets of parameters separately. The basic idea is to use the optimal estimating equation (Crowder, 1986; McCullagh and Nelder, 1989) based on the univariate CL scores to estimate the marginal parameters, and use the pairwise marginal CL score function to estimate the association parameters. Compared to using the pairwise CL alone for estimating both sets of parameters, the combinatorial use of univariate and pairwise CLs was shown to improve the efficiency for the marginal parameters (Kuk, 2007).

Specific to our context, let $l_{ijg(t)|(t-q)}$ denote the univariate marginal-transitional log likelihood contributed by the observation on type g of partner j from household i at time t . The log univariate CL is composed by summing all the univariate marginal-transitional log likelihoods, i.e.

$$\sum_i \sum_t \sum_j \sum_g l_{ijg(t)|(t-q)} = \sum_i \sum_t \sum_j \sum_g (y_{ijg}(t) \log p_{ijg(t)|(t-q)} + (1 - y_{ijg}(t)) \log (1 - p_{ijg(t)|(t-q)})) \quad (6)$$

Additionally, as explained in the next section, instead of using the pairwise marginal CL, we use the weighted conditional pairwise CL $\sum_i \sum_t \sum_j \sum_{j'} \sum_g \sum_{g'} w_{jg,j'g'} l_{ijg(t)|j'g'}$ to estimate the association parameters (denoted by ψ). Let $p_{ijg(t)|j'g'} = P[Y_{ijg}(t) = 1 | y_{ij'g'}(t)]$, then

$$l_{ijg(t)|j'g'} = w_{jg,j'g'} [y_{ijg}(t) \log p_{ijg(t)|j'g'} + (1 - y_{ijg}(t)) \log (1 - p_{ijg(t)|j'g'})] \quad (7)$$

That is, $l_{ijg(t)|j'g'}$ is the conditional pairwise likelihood contributed by the outcome (indexed by jg) conditioning on another concurrently measured outcome (indexed by $j'g'$).

The $l_{ijg(t)|j'g'}$ is a function of the joint probability of $p_{ijg,j'g'}(t) = P[Y_{ijg}(t) = 1, Y_{ij'g'}(t) = 1]$ and the univariate marginal mean of $y_{ij'g'}(t)$ satisfying the marginal-transitional model (4). The $w_{jg,j'g'}$ specifies a weight determining the contribution of the pair $(jg, j'g')$ to the pairwise likelihood. A simple choice is that $w_{jg,j'g'}$ is 1 if $j = j'$ or $g = g'$ (but not both) and 0 otherwise. That is, the pair of binary outcomes contribute to the composite likelihood only when the pair correspond to two different types within a partner (i.e. $j = j'$ and $g \neq g'$) or the same type between the two sex partners (i.e. $j \neq j'$ and $g = g'$). This means outcomes of different types for different partners are assumed independent and thus the corresponding pair do not provide information on the association parameters. More complicated weights may be used to avoid the potential information loss associated with this assumption. In practice, a preliminary data exploration on the magnitudes of the different kinds of associations may help guide the choice of $w_{jg,j'g'}$ (Heagerty and Lele, 1998). Exploration of Rakai data showed that the association between the outcomes of two different types measured on two different partners was much smaller than the within-individual (i.e. inter-type) and between-individual (i.e. intra-couple) associations which were of main interest.

Thus for ease in computation and data management, the simple weight choice was used in the applications.

4.6 Model for the association parameters

The pairwise likelihood function in (7) is indexed by the marginal and joint probabilities of the pair of binary distributions, where the marginal probability is specified by the marginal-transition regression model (4) and parametrized by θ . Although correlation coefficients can be used to parametrize the bivariate joint probability, the odds ratio (OR) is more widely used for measuring association in epidemiology. We therefore parametrize the joint probabilities through association parameters (denoted by ψ) that invite an OR interpretation. Motivated by the alternating logistic regression (ALR) (Carey et al., 1993), the model for ψ is specified as:

$$\text{logit} P[Y_{ijg}(t)=1|y_{ij'g'}(t)] = z_{ijg,j'g'} \begin{pmatrix} \psi_w \\ \psi_b \end{pmatrix} y_{ij'g'}(t) + \log \left(\frac{P_{ijg}(t)|(t-q) - P_{ijg,j'g'}(t)}{1 - P_{ijg}(t)|(t-q) - P_{ij'g'}(t)|(t-q) + P_{ijg,j'g'}(t)} \right) \quad (8)$$

where $z_{ijg,j'g'}$ is the vector of covariates observed for the pair outcomes indexed by $(jg, j'g')$ which includes individual-level covariates (for a pair of two different types within an individual, i.e. $j = j'$ and $g \neq g'$) and couple-level covariates (for a pair of a same type from two sex partners within the same household, i.e. $j \neq j'$ and $g = g'$). If $w_{jg,j'g'} = 0$, i.e. the pair correspond to different types in different partners, then $z_{ijg,j'g'} = \mathbf{0}$, and the relevant data do not contain information for the association parameters ψ .

Here depending on the sources of association, ψ are enforced to have two components: ψ_w is the set of log ORs reflecting within-individual (i.e. inter-type) associations, and ψ_b is the log ORs reflecting between-partner (i.e. intra-couple) associations. Such partition of the association parameters thus connects to the weights we used when constructing the pairwise likelihood in (7). In addition, in the Rakai data with polygamous household, we can further parametrize $\psi_b = (\psi_{bd}^T, \psi_{bid}^T)^T$ where ψ_{bd} corresponds to the association between two partners of a direct sexual relationship (i.e. husband-wife), and ψ_{bid} reflects the association between partners of an indirect sexual relationship (i.e. wife 1-wife 2). As an illustration, in the simple case of $\psi = (\psi_w, \psi_{bd}, \psi_{bid})^T$ (that is, three association parameters characterizing inter-type, between-direct sexual partnerships and between-indirect partners correlations), $z_{ijg;j'g'} = (1, 0, 0)$ for the pair $j = j'$ and $g \neq g'$ (i.e. two types from the same individual), $z_{ijg;j'g'} = (0, 1, 0)$ for the pair $j \neq j', g = g'$ and of different sex (i.e. the same type for the husband and wife), $z_{ijg;j'g'} = (0, 0, 1)$ for the pair $j \neq j', g = g'$ and of same sex (i.e. the same type in two wives from the same household), and $z_{ijg;j'g'} = (0, 0, 0)$ for all other pairs.

5 ESTIMATION AND COMPUTATION

5.1 Estimating equations for the marginal-transitional and association parameters

Let $\mathbf{g}_{ijg}(t)(\theta) = l_{ijg}(t)|(t-q)'\theta$ (i.e., $\mathbf{g}_{ijg}(t)$ is the univariate score function), and let $\mathbf{g}_i(t)$ (of dimension $JGk \times 1$) be the vectorized matrix composed by $(\mathbf{g}_{ijg}(t), j = 1, 2, \dots, J_i; g = 1, 2, \dots, G)$ and $k = \dim \theta$. Following Kuk (2007), we use the optimal linear combinations of the

univariate marginal scores for θ and the score function of the pairwise conditional composite likelihood for ψ .

$$\left(\begin{array}{c} \sum_i \sum_t E(-\frac{\partial \mathbf{g}_i(t)}{\partial \theta} | (t-q))^T \text{cov}(\mathbf{g}_i(t))^{-1} \mathbf{g}_i(t) \\ \sum_i \sum_t \sum_{j \leq j'} \sum_{g \leq g'} w_{jg, j'g'} \frac{\partial l_{ijg(t)|j'g'}}{\partial \psi} \end{array} \right) = 0$$

where the expectation in the first estimating equation is the conditional expectation over the past history as defined in (3), i.e. for component $\mathbf{g}_{ijg}(t)$,

$$E(-\frac{\partial \mathbf{g}_{ijg}(t)}{\partial \theta} | (t-q)) = E(-\frac{\partial \mathbf{g}_{ijg}(t)}{\partial \theta} | \mathcal{A}_{ig}(t-1); \mathcal{B}_{ijg}, \mathcal{C}_{ij(-g)}).$$

It is easy to show that the above estimating equations are equivalent to $U(\theta, \psi) = 0$ where

$$U(\theta, \psi) = \left(\begin{array}{c} \sum_i \sum_t A_{it}^T B_{it}^{-1} C_{it} \\ \sum_i \sum_t D_{it}^T S_{it}^{-1} R_{it} \end{array} \right) = 0 \quad (9)$$

where A_{it} is the $JG \times k$ matrix with columns determined by $p_{ijg(t)|(t-q)'} \theta$, $B_{it} = \text{cov}(\mathbf{y}_i(t))$, C_{it} is the $JG \times 1$ vector with element $(y_{ijg}(t) - p_{ijg(t)|(t-q)})$; and D_{it} is a matrix of dimension $JG(JG-1)/2 \times \text{length}(\psi)$ and with columns determined by $p_{ijg(t)|j'g'} \psi$, S_{it} is the diagonal matrix with diagonal element $p_{ijg(t)|j'g'}(1 - p_{ijg(t)|j'g'})$ for $j = j'$ and $g = g'$, and R_{it} is the $JG(JG-1)/2 \times 1$ vector with element $(y_{ijg}(t) - p_{ijg(t)|j'g'})$.

Let $(\hat{\theta}, \hat{\psi})$ be the solutions from the estimating equations in (9). The large sample properties of $(\hat{\theta}, \hat{\psi})$ are presented in Supplementary Materials-S1. They can be derived by applying the asymptotic properties of composite likelihood estimates (Varin et al., 2011) with the additional use of the asymptotic properties of martingales.

Note that the first estimating equation for the marginal-transitional parameters θ in (9) is similar to the optimal estimating equation given by (McCullagh and Nelder, 1989, section 9.4.2) for an autoregressive process, with the caveat that the covariance matrix is not diagonal and is dependent on the association parameters to be estimated from the second equation in (9), thus extending to our specific context where there are multiple correlated Markovian processes.

It was also pointed out in Kuk (2007) that the alternating logistic regression originally proposed in the context of generalized estimating equations can be seen as a special situation of this modified pairwise pseudolikelihood method. Estimation of the association parameters in ALR essentially uses the score equation of the composite likelihood composed by pairwise conditional likelihoods (Kuk, 2007, 2004).

5.2 Computation

With a close connection to the ALR model, one advantage of our proposed analytical method is that computation can be performed with commonly used softwares that have routines for ALR. The main challenge in computation thus is data management: for the marginal-transition model, a dataset is needed where a subject's own history and the

partner's past testing result (as specified in (4)) have been manipulated into columns parallel to the response column of current testing result; for the association model (9), programming effort is needed to construct the design matrix. To leave flexibility for examining different specifications of association parameters, our suggestion is to first construct a dataset where each row contains covariate information for a pair of observations (indexed by (j, g) and (j', g')) belonging to the same household and the same visit. The design matrix for the association model then can be built depending on the covariate values and specification of the association parameters of interest. SAS PROC GENMOD was used in our applications, and more details about the required data formats are illustrated in Supplementary Materials-S2. Another computation tool is the "orth" package in R (By et al., 2014), and data of similar formats need to be prepared for using this package.

6 ANALYSIS OF THE HPV COUPLE DATA FROM THE RAKAI MALE CIRCUMCISION TRIAL

6.1 Model building strategy for the marginal-transitional model

HPV infection is often transient with a median duration of 6-months (Morse et al., 2003). As the Rakai HPV couple data were collected at one-year intervals and there were only two follow-up visits, we used a first order transition model (i.e. $q = 1$) where the past history included a subject's own testing result and his/her partners' testing result, that is, $H_1(\cdot)$ and $H_2(\cdot)$ in model (5) were taken to be 0. With this simplification, the marginal-transition regression model can still be cumbersome when including covariates (e.g. gender, circumcision status). Thus to answer each of the aforementioned research questions (Section 3.2), we adopted a "step-down" model fitting procedure (Kleinbaum and Klein, 2002, Chapter 6) that started by including all relevant terms and up to their three-way interactions. Interaction terms with a significance level > 0.1 (based on Wald tests) were removed from the final model. Note however, the interaction term between a subject's own and his/her partners' prior testing results is always retained, leading to a saturated model for the transition probability $P(Y_{ijg}(t) = 1 | \{y_{ijg}(t-1), I(\sum_{j'} y_{ij'g}(t-1) > 0)\})$. Moreover, the interactions between a covariate and a subject's own or the partners' infection status at $t-1$ were always retained in the model to allow different effects of the covariate on the outcomes of transmission and persistence.

We used data of all the 640 couples to answer the question of effect of randomization to MC on preventing HPV transmission, allowing assessment of how the HPV infection status of one female partner affected that of another female partner if they had a sexual relationship with a same man (i.e. the association between indirect partnerships). However, for the other question on comparing male-to-female and female-to-male transmission risks, we used the subset of 412 couples where both of the partners reported only having one sex partner in the past year at the two follow-up visits. This is to minimize the potential misclassification caused by the unobserved HPV status of sex partners not enrolled in the Rakai MC study.

6.2 Application 1 - effect of male circumcision on HPV transmission and persistence in men and women

Following the model building strategy described in section 6.1 and using the logit link in the marginal-transition model (5) for polygamous households, the selected model is:

$$\begin{aligned} \text{logit} \left(p_{ijg(t)|(t-q)} \right) = & \mu + \beta_1 y_{ijg}(t-1) + \beta_2 I \left(\sum_{j' \neq j} y_{ij'g}(t-1) > 0 \right) \\ & + \beta_{12} y_{ijg}(t-1) \cdot I \left(\sum_{j' \neq j} y_{ij'g}(t-1) > 0 \right) \\ & + \alpha_1 x_{1ij}(t) + \alpha_2 x_{2ij}(t) + \alpha_{12} x_{1ij}(t) \cdot x_{2ij}(t) \\ & + \alpha_{1\beta_1} x_{1ij}(t) \cdot y_{ijg}(t-1) + \alpha_{1\beta_2} x_{1ij}(t) \cdot I \left(\sum_{j' \neq j} y_{ij'g}(t-1) > 0 \right) \\ & + \alpha_{2\beta_1} x_{2ij}(t) \cdot y_{ijg}(t-1) + \alpha_{2\beta_2} x_{2ij}(t) \cdot I \left(\sum_{j' \neq j} y_{ij'g}(t-1) > 0 \right) \end{aligned} \quad (10)$$

where x_{1ij} refers to the circumcision status of the male partner in household i (1="circumcised") and x_{2ij} is gender of individual j in household i (1="female"). In addition, 8 association parameters ($\psi_1, \psi_1, \dots, \psi_8$) were specified in the association model (9) to quantify the different kinds of associations by gender and circumcision status. The parameter estimates can be found in Supplementary Materials-S3.

Based on the coefficient estimates in Model (10), Table 2 summarizes the estimated effects of MC (presented as odds ratios) on male-to-female and female-to-male transmission, new HPV detection from an unidentified source, and persistence of a pre-existing infection. If a man was circumcised, his odds of acquiring an HPV infection from his partner (i.e. female-to-male transmission) was estimated to be halved (95% confidence interval [CI] 0.28 to 0.89), and the odds of new HPV detection from an unknown source was reduced to 0.45 (95% CI 0.30 to 0.67) compared to uncircumcised control arm men. For a woman, the point estimates showed that having a circumcised male partner reduced her odds of acquiring an infection from the partner (i.e. male-to-female transmission) by 20%, and decreased her odds of having a newly detected HPV from an unknown source by 28%. As it is rare for Rakai women to engage in non-marital sexual relationships, the unknown source of HPV could be their concurrently enrolled husbands who may have acquired the infections during the year between $t-1$ and t and subsequently transmitted them to the women. In this case, the similar magnitude of effects of MC on transmissions from the male partner and unknown sources would suggest that MC may confer some degree of protection against HPV acquisition in women. The model based results on HPV transmission (Table 2) were also compared to the naive OR estimates obtained by summarizing the number of transmissions of discordant infections (Table 1): the ORs were in the same direction for both male-to-female and female-to-male transmissions. By adopting a modeling framework, the model-based estimates gained higher efficiency (Agresti, 2002, Chapter 6) compared to the naive estimates which were based on the sample proportions (Table 2) and assumed independence between the discordant infections.

However, it is important to note that the new "detections", including the transmissions (as operationally defined) from the known partners, are identified based on the observed DNA testing results and may be a mixture of truly new acquisitions from the sex partner and

reactivations of latent infections (Xi et al., 2011; Winer et al., 2014). Over 98% of the MC trial participants had been sexually active for more than 2 years before the trial enrollment. Thus, a seemingly new acquisition may be equally likely a reactivation of an earlier infection that became latent and undetectable. The reduction in HPV prevalence in the circumcised men hence could also be a result of MC's effect on reducing the risk of reactivation, possibly through removing the foreskin tissue where the virus may have persisted with limited viral gene expression (i.e. a latent state) (Doorbar, 2013). For women, it is unknown whether having sex with a circumcised partner could induce less "mechanical irritation" or "extracellular influences" (Doorbar, 2013). If so, MC may also play a role in reducing the risk of reactivation in the female partners as Doorbar (2013) suggested that along with host immune regression, inflammation could also contribute to reactivation. Given our current imperfect knowledge about the natural history of HPV infection and the lack of molecular tool to distinguish new infections from reactivations (Xi et al., 2011; Gravitt, 2012; Winer et al., 2014), the mechanisms (reducing transmission, reactivation, or both) through which MC act on reducing HPV detection in men and their female partners cannot be definitely determined.

Regarding persistence of pre-existing infections, circumcision significantly decreased persistence of a pre-existing infection in men, irrespective of the female partner's infection status at $t - 1$; but it did not significantly reduce the odds of persistence in women, irrespective of whether the male partner had the infection at $t - 1$. In summary, MC showed a strong and direct protective effect in men by reducing their odds of new HPV detection (including transmission from the known partners) and persistence of pre-existing infections. Women benefited indirectly and to a lesser extent from MC of their male partners possibly through reduced exposure to penile HPV during sex.

We also assessed the strengths of associations between the infection status of different types within an individual, and between direct and indirect partnerships on a same type. Estimates of the parameters characterizing such associations are presented in Table 3. In both men and women, carrying one type of HPV significantly increased the odds of the subject being infected with another type, and circumcision status did not seem to have much impact on such association. For example, for a circumcised man, if he was observed to carry one type of HPV, his odds of being infected with another type was estimated to increase by a factor of 3.9 (95% CI 1.9 to 8.1). In addition, there was a strong association between partners of direct sexual relationship: in couples where the men were uncircumcised, one partner's harboring an HPV substantially increased the odds of the other partner having the same type of HPV (estimated OR=14.2, 95% CI 8.6 to 23.3), and such association was reduced among couples with circumcised men (estimated OR=7.6, 95% CI 3.1 to 18.5). The difference between circumcised and uncircumcised couples in the degree of between-partner associations in fact is consistent with our earlier observations on MC's protective effects on reducing the HPV burden in both men and women (Table 2). Significant association was also observed between wives of a common husband who was uncircumcised (i.e. between indirect partnerships), but such association was not significant when the husband was circumcised (estimated OR=0.28, 95% CI 0.02 to 4.3). It has been shown that MC reduced HPV viral load in circumcised HPV-positive men and their HPV-positive female partners

(Davis et al., 2013; Wilson et al., 2014). Thus possibly by removing the foreskin (a reservoir for viral replication) and consequently reducing the amount of penile HPV virus, MC could reduce cross-infections among the multiple sex partners. It should also be noted that since the associations between indirect partnerships could only be estimated from polygamous households whereas the associations between direct partnerships were estimated from all households, confounding factors may exist that could explain the different levels of associations between indirect and direct partnerships among households with circumcised men. However, the levels of associations between direct and indirect partnerships were similar among households with uncircumcised men, thus such confounding, if existed, did not seem to play a role among these households.

6.3 Application 2 - associations of gender on HPV transmission and persistence

The subset of 412 couples reporting no non-marital partners in the past year was used to assess the differences in HPV transmission and persistence between men and women. Because the MC status of men is significantly related to male risk of HPV infection, the comparisons between genders are presented by MC status in Table 4. The final model selected similar model terms as those in (10) and thus is not repeated here. There were 6 association parameters ($\psi_1, \psi_1, \dots, \psi_6$) corresponding to the inter-type association by gender and MC status and the intra-couple association by gender. The coefficient estimates are presented in Supplementary Materials-S4.

Among couples with circumcised men (Table 4), the estimated odds of being infected with the virus carried by the partner at the previous visit was over 50% higher in women than men. This was not statistically significant, possibly due to the small numbers of transmission events observed. The odds of newly detected HPV from unknown sources was significantly higher in women than men (OR=1.93, 95% CI 1.3 to 3.0). If both partners of the couples truly had no extra-marital partners, the unknown sources were possibly the other partner in the partnership, in which case taking the two effects together, women seemed to have a higher risk of acquiring HPV than men if the men were circumcised; but such a gender difference was not seen in couples with uncircumcised men. Alternatively, the new detection from an unknown source could also represent reactivations of latent infections, in which case the odds of reactivation was estimated higher in women than men, particularly when the men were circumcised.

For persistence, irrespective of the partner's infection status at $t - 1$, women always showed higher odds than men, especially among couples with circumcised men. This is consistent with the belief that the duration of HPV infection may be shorter in men than in women (Rowhani-Rahbar et al., 2009). Cervical cancer is the only HPV-related cancer that is among the eight most common cancer types worldwide. As persistence of oncogenic HPV infection is known as a necessary step leading to anogenital cancers (Cogliano et al., 2005), our observation of the higher risk of HPV persistence in women (irrespective of the MC status of their male sex partners) is consistent with the observations of a higher HPV-associated cancer burden in women.

Estimates for the within-individual and between-direct partnerships associations are very similar to those obtained in Application 1 (Table 3) and are thus omitted here. In fact, the

coefficient estimates for all model terms are comparable between Application 1 which used data from all couples (Table S-3) and Application 2 (Table S-4) which focused on data of the subset of couples reporting no non-marital partners. Additionally, to check on the robustness of the numerical results obtained from the asymptotic distribution, the non-parametric bootstrap confidence intervals were obtained and are very comparable to the results from applying the large sample properties. More details about the bootstrap analysis and results can be found in the Supplementary Materials-S5.

7 Discussion

In multivariate outcome data repeatedly measured on clusters, existing approaches for correlated data are not applicable due to the possible simultaneous presence of correlations of distinct nature. In the context of HPV transmission dynamics within couples, data on multiple HPV types are often available from each partner, and the temporal correlation can translate to the scientific interest on the event of “transmission”. Motivated by the operational definition of HPV transmission, we proposed a hybrid modeling strategy that combinatorially uses the Markovian transition model and pairwise conditional likelihood method. Although the Markov transition model has received relatively less attention compared to marginal or random effects models for the analysis of correlated binary data, it offers a natural tool when the scientific interest itself focuses on an event defined by the past history (such as incidence, recurrence, etc.). This proposed method needs nontrivial data manipulation, but existing software can be used directly for estimation, which may facilitate its usage in practice.

HPV is a common sexually transmitted infection, but challenges and difficulties exist when estimating HPV transmission (Rowhani-Rahbar et al., 2009). The limitations inherent in the data generated from HPV DNA assays, along with our imperfect knowledge about the natural history of HPV, can affect the scientific inference drawn from statistical analysis. Our proposed model followed the commonly used operational definition of “transmission” which is based on observed testing results. Such testing results however may be subject to misclassification: a positive result may indicate contamination from a sex partner rather than a true, established infection (Burchell et al., 2011), and a negative result may be due to an infection being in latency or having undetectable viral shedding (Gravitt, 2011). The assay cannot differentiate reactivation of a latent infection from a new infection transmitted from the partner. Also, an infection may not be immediately detectable, and thus there could be a lag between the time of infection and the time of detection. These uncertainties consequently may cause bias in the inference drawn from data analysis. Introducing a hidden state to represent the true infection status as in Bureau et al. (2003), may be considered to account for misclassification. Misclassification rates may also be treated as pre-specified parameters and incorporated in the pairwise likelihood function as in Kang and Lagakos (2007). Additionally, as pointed out by a reviewer, without viral sequencing data, there is uncertainty in ascribing a newly-detected infection from exposure to the particular partner enrolled in the study. While one can limit the analysis to data from couples reporting no non-marital sexual relationships, uncertainty remains as self-reported sexual behaviors are often subject to reporting bias. Another limitation of our proposed modeling framework is that for the outcome of persistence, the method models the odds of persistence of a pre-

existing infection at the subsequent follow-up visit, hence, it does not consider how long that infection has existed. Such a definition is unsatisfactory when the scientific interest is the duration of infection and the study design has relatively short follow-up intervals. The Markov or semi-Markov models proposed by Bureau et al. (2003), Kang and Lagakos (2007) and Mitchell et al. (2011) offer a useful modeling framework for studying the duration of persistence of an individual HPV type. If the interest is studying risk factors for HPV persistence or comparing the persistence of different types, a simultaneous use of data on multiple types will be preferred. Research is needed to accommodate such multivariate data, and the pairwise composite likelihood approach may be useful.

The marginal-transition model proposed in (4) essentially specifies a time-homogeneous Markov model. If a couple study has a long-term follow-up and the predictor variables of interest are expected to have a time-varying effect on transmission (such as a HPV vaccine that may confer increasing protection over time), a time-heterogeneous model may be used by specifying time-dependent coefficients in the marginal-transition model. Additionally, while the length of the follow-up interval is often fixed by design in longitudinal HPV studies, unequal intervals may arise due to logistic reasons or study participants missing scheduled visits. A time-heterogeneous model may be used to accommodate the unequal sampling intervals where the effects of predictor variables may be allowed to differ by interval length. Model (4) assumes order- q dependence on the past history. Recent studies suggest previous infection may render some immune protection against redetection of the same type of infection (Moscicki et al., 2013; Lin et al., 2013). This may be examined by extending the order- q dependence assumption in Model (4) and modeling a subject's current infection status as a function of all past testing results observed during the study (i.e. a non-zero $H_1(\cdot)$). Moreover, our model handles data on multiple HPV types simultaneously, allowing the assessment of possible cross-type immunity from prior infections. Although earlier studies suggested low cross-type immunity (Trottier and Franco, 2006; Burchell et al., 2006b), this question may be of new interest given the recent concern of HPV type replacement (Tota et al., 2013) with the use of HPV vaccines (which only covers HPV-6, 11, 16 and 18). Another potential use of the method is to study auto-inoculation from cohort studies with assay data from multiple anatomic sites (e.g. penis, scrotum, hand, etc.) from the same individual at each visit. The dynamic event of interest will be the transmission of HPV from one site to another site within a person (i.e. auto-inoculation), and the different anatomic sites take the role of the different "partners" in our proposed method.

Missing data often occur in practice. In the Rakai MC trial data, missing HPV results arose when samples were unavailable due to couples lost to follow-up, or when samples were available but unamplifiable due to insufficient cellular materials. The main reason for sample unavailability was an unexpected storage outage of HPV swabs which affected both arms and both genders, and thus this missing data mechanism can be considered as MCAR (missing completely at random, Little and Rubin (2002)). The missingness due to insufficient cells however highly depended on gender and circumcision status of the man: 99% of the female swabs were amplifiable, whereas only 83% and 66% of swabs from uncircumcised and circumcised men respectively could be amplified. Since the amplification of the human β -globin gene in the sample was used as an internal control for sample adequacy and this

gene is not confounded with HPV DNA, the sample's amplifiability should not depend on the presence/absence of the virus. Therefore, this source of missing data may be considered as MAR (missing at random, Little and Rubin (2002)) where missingness depends on gender and circumcision status. In our application analyses, because MC status and gender were the predictor variables of primary interest when assessing the effect of MC on female and male HPV, they were always included in both the marginal-transition model and the association model (i.e. different sets of association parameters by gender and MC). Thus, applying our model on the observed data is scientifically relevant and expected to result in statistically proper estimates.

An order-one transitional model in general may not be sufficient (Bureau et al., 2003; Mitchell et al., 2011); however, we were limited to using the first order transitional models given the small number of follow-up visits in the Rakai data (second order models were also explored and the details can be found in the Supplemental Materials-S6). The majority of the partnerships had been formed for more than 2 years at the time of trial enrollment, and it is impossible to distinguish a truly new acquisition from a reactivation (Xi et al., 2011). However, as other MC trials in Africa did not enroll female partners, the Rakai data provided a unique opportunity to examine MC's effect on HPV infection in heterosexual couples. Our analyses on the efficacy of male circumcision on oncogenic HPV transmission and persistence and on between-gender comparisons may inform future mathematical modeling studies on the cost-effectiveness of MC, HPV vaccination and other preventive strategies. The strong effect of MC on preventing transmission (and/or reactivation) and persistence suggests that in a modeling study, the prevalence of MC in the target population is an important population characteristic to consider. A few other prospective couple studies of HPV are in progress in Canada (the HITCH cohort study) (Burchell et al., 2010b,a) and the US (Hernandez et al., 2008; Widdice et al., 2010; Nyitray et al., 2014). These couple studies are generating rich arrays of HPV data, and our proposed model may offer a tool for studying the dynamics and characteristics of HPV transmission in heterosexual couples. It is particularly useful for identifying risk factors for HPV transmission (e.g., condom use, frequency of sexual contact), comparing transmission risks of different HPV types, and characterizing inter-type and intra-couple associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Agresti, A. Categorical data analysis. Wiley-Interscience; 2002.
- Bonney G. Logistic regression for dependent binary observations. *Biometrics*. 1987; 43:951–973. [PubMed: 3427178]
- Breslow N, Clayton D. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993; 88:9–25.
- Burchell A, Coutlee F, Tellier P, Hanley J, Franco E. Genital transmission of human papillomavirus in recently formed heterosexual couples. *Journal of Infectious Diseases*. 2011; 204(11):1723–1729. [PubMed: 21984739]
- Burchell A, Richardson H, Mahmud S, Trottier H, Tellier P, Hanley JFC, Franco E. Modeling the sexual transmissibility of human papillomavirus infection using stochastic computer simulation and

- empirical data from a cohort study of young women in montreal, canada. *American Journal of Epidemiology*. 2006a; 163(6):534–543. [PubMed: 16421235]
- Burchell AN, Tellier PP, Hanley J, Coutlee F, Franco EL. Influence of partner's infection status on prevalent human papillomavirus among persons with a new sex partner. *Sexually Transmitted Diseases*. 2010a; 37(1):34–40. [PubMed: 19704391]
- Burchell AN, Tellier PP, Hanley J, Coutlee F, Franco LE. Human papillomavirus infections among couples in new sexual relationships. *Epidemiology*. 2010b; 21(1):31–37. [PubMed: 19907332]
- Burchell AN, Winer RL, de Sanjose S, Franco EL. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. *Vaccine*. 2006b; 24(Suppl 3):S3/52–61. [PubMed: 16950018]
- Bureau A, Shiboski S, Hughes J. Applications of continuous time hidden markov models to the study of misclassified disease outcomes. *Statistics in Medicine*. 2003; 22(3):441–462.
- By K, Qaqish B, Preisser J, Perin J, Zink R. ORTH: R and SAS software for regression models of correlated binary data based on orthogonalized residuals and alternating logistic regressions. *Computer Methods and Programs in Biomedicine*. 2014; 113(2):557–568. [PubMed: 24286728]
- Carey V, Zeger S, Diggle P. Modeling multivariate binary data with alternating logistic regressions. *Biometrika*. 1993; 80:517–526.
- Cherpes TL, Melan MA, Kant JA, Cosentino LA, Meyn LA, Hillier SL. Genital tract shedding of herpes simplex virus type 2 in women: effects of hormonal contraception, bacterial vaginosis, and vaginal group B *Streptococcus* colonization. *Clinical Infectious Diseases*. 2005; 40(10):1422–1428. [PubMed: 15844064]
- Chib S, Jeliazkov I. Inference in semiparametric dynamic models for binary longitudinal data. *Journal of the American Statistical Association*. 2006; 101(474):685–700.
- Cogliano V, Baan R, Straif K, Grosse Y, Secretan B, Ghissassi F. WHO International Agency for Research on Cancer. Carcinogenicity of human papillomaviruses. *The Lancet Oncology*. 2005; 6(4):204. [PubMed: 15830458]
- Crowder M. On consistency and inconsistency of estimating equations. *Econometric Theory*. 1986; 2(3):305–330.
- Davis M, Gray R, Grabowski M, Serwadda D, Kigozi G, Gravitt P, Nalugoda F, Watya S, Wawer M, Quinn T, Tobian A. Male circumcision decreases high-risk human papillomavirus viral load in female partners: a randomized trial in Rakai, Uganda. *International Journal of Cancer*. 2013; 133:1247–1252.
- Doorbar J. Latent papillomavirus infections and their regulation. *Current Opinion in Virology*. 2013; 3:416–421. [PubMed: 23816390]
- Efron, B.; Tibshirani, B. *An introduction to the Bootstrap*. Chapman & Hall/CRC; Boca Raton: 1993.
- Fahrmeir L, Kaufmann H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*. 1985; 13(1):342–368.
- Firth D. Discussion of the paper by Liang, Zeger and Qaoish. *Journal of the Royal Statistical Society Series B*. 1992; 45:28–9.
- FitzGerald E, Knuiman M. Interpretation of regressive logistic regression coefficients in analyses of familial data. *Biometrics*. 1998; 54:909–920.
- Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G. *Longitudinal data analysis*. CRC Press; Boca Raton: 2009.
- Franco E, Bosch F, Cuzick J, Schiller J, Garnett G, Meheus A, Wright T. Chapter 29: Knowledge gaps and priorities for research on prevention of HPV infection and cervical cancer. *Vaccine*. 2006; 24(Suppl 3):S3/242–9.
- Geys H, Molenberghs G, Ryan L. Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*. 1999; 94(447):734–745.
- Giuliano A, Lu B, Nielson C, Flores R, Papenfuss M, Lee J, Abrahamsen M, Harris R. Age-specific prevalence, incidence, and duration of human papillo-mavirus infections in a cohort of 290 US men. *Journal of Infectious Diseases*. 2008; 198(6):827–835. [PubMed: 18657037]
- Gravitt P. The known unknowns of HPV natural history. *The Journal of Clinical Investigation*. 2011; 121(12):4593–4599. [PubMed: 22133884]

- Gravitt P. Evidence and impact of human papillomavirus latency. *The Open Virology Journal*. 2012; 6(Suppl 2):198–203. [PubMed: 23341855]
- Gray R, Kigozi G, Serwadda D, Makumbi F, Watya S, Nalugoda F, Kiwanuka N, Moulton L, Chaudhary M, Chen M, Sewankambo N, Wabwire-Mangen F, Bacon M, Williams C, Opendi P, Reynolds S, Laeyendecker O, Quinn T, Wawer M. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet*. 2007; 369(9562):657–666. [PubMed: 17321311]
- Halfon P, Benmoura D, Agostini A, Khiri H, Penaranda G, Martineau A, Blanc B. Evaluation of the clinical performance of the Abbott RealTime High-Risk HPV for carcinogenic HPV detection. *Journal of Clinical Virology*. 2010; 48(4):246–50. [PubMed: 20541966]
- Heagerty P, Lele S. A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*. 1998; 93(443):1099–1111.
- Hernandez BY, Wilkens LR, Zhu X, Thompson P, McDuffe K, Shvetsov YB, Kamemoto LE, Killeen J, Ning L, Goodman MT. Transmission of human papillomavirus in heterosexual couples. *Emerging Infectious Diseases*. 2008; 14(6):888–894. [PubMed: 18507898]
- Heyde, C. *Quasi-Likelihood And Its Application*. Springer; New York: 1997.
- Hjort N, Varin C. ML, PL, QL in Markov Chain Models. *Scandinavian Journal of Statistics*. 2007; 35(1):64–82.
- Kang M, Lagakos S. Evaluating the role of human papillomavirus vaccine in cervical cancer prevention. *Statistical Methods in Medical Research*. 2004; 13:139–155. [PubMed: 15068258]
- Kang M, Lagakos S. Statistical methods for panel data from a semi-Markov process, with application to HPV. *Statistics in Medicine*. 2007; 8(2):252–264.
- Kim J, Goldie S. Cost effectiveness analysis of including boys in a human papillomavirus vaccination programme in the United States. *BMJ*. 2009; 339:b3884. [PubMed: 19815582]
- Kleinbaum, D.; Klein, M. *Logistic Regression A self-Learning Text*. 2. Springer; 2002.
- Kong X, Gray R, Archer K, Moulton L, Wang M. Parametric Frailty Models for Clustered Data with Arbitrary Censoring: Application to Effect of Male Circumcision on HPV Clearance. *BMC Medical Research Methodology*. 2010a; 10:40. [PubMed: 20459614]
- Kong X, Gray R, Moulton L, Wawer M, Wang M. A modeling framework for the analysis of HPV incidence and persistence: a semi-parametric approach for clustered binary longitudinal data analysis. *Statistics in Medicine*. 2010b; 29(28):2880–2889.
- Kuk A. Permutation invariance of alternating logistic regression for multivariate binary data. *Bimetrika*. 2004; 91(3):758–761.
- Kuk A. A hybrid pairwise likelihood method. *Bimetrika*. 2007; 94(4):939–952.
- Kuk A, Nott D. A pairwise likelihood approach to analyzing correlated binary data. *Statistics & Probability Letters*. 2000; 47(4):329–335.
- Le Cessie S, Van Houwelingen J. Logistic regression for correlated binary data. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. 1994; 43(1):95–108.
- Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- Liang K, Zeger S. A class of logistic regression models for multivariate binary time series. *Journal of the American Statistical Association*. 1989; 84:447–451.
- Lin S, Ghosh A, Porras C, Markt S, Rodriguez A, Schiffman M, Wacholder S, Kemp T, Pinto L, Gonzalez P, Wentzensen N, Esser M, Matys K, Meuree A, Quint W, van Doorn L, Herrero R, Hildesheim A, Safaeian M. Group CRVT. HPV16 seropositivity and subsequent HPV16 infection risk in a naturally infected population: comparison of serological assays. *PLoS One*. 2013; 8(1)
- Lindsay B. Composite likelihood methods. *Contemporary Mathematics*. 1988; 80:221–239.
- Little, R.; Rubin, D. *Statistical Analysis with missing data*. Wiley, Hoboken; New Jersey: 2002.
- McCullagh, P.; Nelder, J. *Generalized linear models*. Chapman and Hall/CRC; Boca Raton: 1989.
- Mitchell C, Hudgens M, King C, Cu-Uvin S, Lo Y, Rompalo A, Sobel J, Smith J. Discrete-time semi-markov modeling of human papillomavirus persistence. *Statistics in Medicine*. 2011; 30(17): 2160–2170.
- Molenberghs, G.; Verbeke, G. *Models for discrete longitudinal data*. Springer; New York: 2005.

- Morse, S.; Ballard, R.; Holmes, K.; Moreland, A. *Atlas of Sexually Transmitted Diseases and AIDS*. 3. Mosby, Edinburgh, London, New York, Oxford, Philadelphia, St Louis, Sydney, Toronto: 2003.
- Moscicki A, Ma Y, Farhat S, Darragh T, Pawlita M, Galloway D, Shiboski S. Redetection of Cervical Human Papillomavirus Type 16 (HPV16) in Women With a History of HPV16. *The Journal of Infectious Diseases*. 2013; 208:403–412. [PubMed: 23599313]
- Nyitray A, Lin H, Fulp W, Chang, Menezes L, Lu B, Abrahamsen M, Papenfuss M, Gage C, Galindo C, Giuliano A. The role of monogamy and duration of heterosexual relationships in human papillomavirus transmission. *The Journal of Infectious Diseases*. 2014; 209:1007–1015. [PubMed: 24253288]
- O'Brien L, Fitzmaurice G. Analysis of longitudinal multiple-source binary data using generalized estimating equations. *Applied Statistics*. 2004; 53(1):177–193.
- Palefsky JM. Human papillomavirus-related disease in men: not just a women's issue. *The Journal of Adolescent Health*. 2010; 46(4 Suppl):S12–19. [PubMed: 20307839]
- Parner E. A composite likelihood approach to multivariate survival data. *Scandinavian Journal of Statistics*. 2001; 28(2):295–302.
- Rowhani-Rahbar A, Hughes J, Koutsky L. Difficulties in estimating the male-to-female sexual transmissibility of human papillomavirus infection. *Sexually Transmitted Disease*. 2009; 36(4): 261–263.
- Ten Have T, Morabia A. Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometric*. 1999; 55:85–93.
- Tota J, Ramanakumar A, Jiang M, Dillner J, Walter S, Kaufman J, Coutle F, Villa L, Franco E. Epidemiologic approaches to evaluating the potential for human papillomavirus type replacement postvaccination. *American Journal of Epidemiology*. 2013; 178(4):625–634. [PubMed: 23660798]
- Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. *Vaccine*. 2006; 24(Suppl 1):S1–15. [PubMed: 16406226]
- Varin C, Reid N, Firth D. An overview of composite likelihood methods. *Statistica Sinica*. 2011; 21:5–42.
- Varin, Ca. On composite marginal likelihoods. *ASTA Advances in Statistical Analysis*. 2008; 92(1):1–28.
- WHO International Agency for Research on Cancer. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Human Papillomaviruses*. IARC Press; Lyon, France: 2007.
- Widdice LE, Breland DJ, Jonte J, Farhat S, Ma Y, Leonard AC, Moscicki AB. Human papillomavirus concordance in heterosexual couples. *The Journal of Adolescent Health*. 2010; 47(2):151–159. [PubMed: 20638007]
- Wilson L, Gravitt P, Tobian A, Kigozi G, Serwadda D, Nalugoda F, Watya S, Wawer M, Gray R. Male circumcision reduces penile high-risk human papillomavirus viral load in a randomised clinical trial in Rakai, Uganda. *International Journal of Cancer*. 2014; 134:1889–1898.
- Winer R, Xi L, Shen Z, Stern J, Newman L, Feng Q, Hughes J, Koutsky L. Viral load and short-term natural history of type-specific oncogenic human papillomavirus infections in a high-risk cohort of midadult women. *International Journal of Cancer*. 2014; 134:1889–1898.
- Xi L, Hughes J, Castle P, Edelstein Z, Wang C, Galloway D, Koutsky L, Kiviat N, Schiffman M. Viral load in the natural history of human papillomavirus type 16 infection: a nested case-control study. *The Journal of Infectious Diseases*. 2011; 203:1425–1433. [PubMed: 21415020]
- Zhao Y, Joe H. Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*. 2005; 33(3):335–356.

(a) Intervention arm

(b) Control arm

Figure 1.
Prevalence of HPV infection in men and women by randomization arm and visit. (GT: HPV type)

Table 1

Number of male-to-female and female-to-male transmission events by circumcision status of the husband.
(MC: male circumcision. OR: odds ratio. CI: confidence interval)

MC status	Total No. of discordant infections *	No. transmitted [†]	OR MC Yes vs. No.	95% CI
Female-to-Male Transmission				
Yes	114	9	0.62	(0.27,1.41)
No	173	21		
Male-to-Female Transmission				
Yes	102	8	0.49	(0.21,1.17)
No	135	20		

* The "Total No. of discordant infections" for female-to-male transmission is the sum of the number of discordant infections where the female partner was tested positive and the male partner was tested negative at baseline and the number of such discordant infections at the first follow-up visit. The "Total No. of discordant infections" for male-to-female transmission is the sum of the number of discordant infections where the male partner was tested positive and the female partner was tested negative at baseline and the number of such discordant infections at the first follow-up visit.

[†]The "No. transmitted" refers to the number of discordant infections where the initially negative partner became positive at the next visit.

Table 2

Estimated effects of male circumcision on HPV transmission, new detection from an unknown source, and persistence. (OR: odds ratio. MC: male circumcision)

MC status of the husband (yes vs. no).	Effect in men		Effect in women	
	Estimated OR	95% CI	Estimated OR	95% CI
HPV transmission	$\exp(a_1 + a_{1\beta_2})$ 0.50	(0.28, 0.89)	$\exp(a_1 + a_{1\beta_2} + a_{12})$ 0.80	(0.47, 1.38)
New detection from an unknown source	$\exp a_1$ 0.45	(0.30, 0.67)	$\exp(a_1 + a_{12})$ 0.72	(0.52, 1.00)
Persistence when the partner did not carry the infection at $t - 1$	$\exp(a_1 + a_{1\beta_1})$ 0.51	(0.30, 0.88)	$\exp(a_1 + a_{1\beta_1} + a_{12})$ 0.83	(0.54, 1.27)
Persistence when the partner also carried the infection at $t - 1$	$\exp(a_1 + a_{1\beta_1} + a_{1\beta_2})$ 0.57	(0.32, 1.03)	$\exp(a_1 + a_{1\beta_1} + a_{1\beta_2} + a_{12})$ 0.92	(0.56, 1.53)

Table 3

Associations within an individual and between partners (OR: odds ratio.)

MC status of the husband	Circumcised		Uncircumcised	
	Estimated OR	95% CI	Estimated OR	95% CI
Within-individual i.e. between-type association in men	exp ψ_1		exp ψ_2	
	3.90	(1.88, 8.06)	2.70	(1.71, 4.25)
Within-individual i.e. between-type association in women	exp ψ_3		exp ψ_4	
	3.07	(1.69, 5.56)	4.93	(2.78, 8.72)
Between-direct partnerships	exp ψ_5		exp ψ_6	
	7.57	(3.09, 18.52)	14.19	(8.64, 23.32)
Between-indirect partnerships	exp ψ_7		exp ψ_8	
	0.28	(0.02, 4.31)	14.08	(2.15, 92.33)

Table 4

Associations of gender on HPV transmission, new detection with unknown sources, and persistence. (OR: odds ratio. MC: male circumcision. F: female. M: male.)

Circumcised	Estimated OR	95% CI
M-to-F vs. F-to-M transmission	$\exp(a_2 + a_{12} + a_{2\beta_2})$	
	1.53	(0.76, 3.08)
New detection with unknown sources F vs. M	$\exp(a_2 + a_{12})$	
	1.93	(1.26, 2.95)
Persistence when partner not carrying the infection at $t - 1$: F vs. M	$\exp(a_2 + a_{12} + a_{2\beta_1})$	
	3.57	(1.96, 6.5)
Persistence when partner carrying the infection at $t - 1$: F vs. M	$\exp(a_2 + a_{12} + a_{2\beta_1} + a_{2\beta_2})$	
	2.83	(1.68, 4.77)
Uncircumcised		
M-to-F vs. F-to-M transmission	$\exp(a_2 + a_{2\beta_2})$	
	1.03	(0.56, 1.88)
New detection with unknown sources F vs. M	$\exp(a_2)$	
	1.30	(0.95, 1.77)
Persistence when partner not carrying the infection at $t - 1$: F vs. M	$\exp(a_2 + a_{2\beta_1})$	
	2.40	(1.38, 4.17)
Persistence when partner carrying the infection at $t - 1$: F vs. M	$\exp(a_2 + a_{2\beta_1} + a_{2\beta_2})$	
	1.91	(1.26, 2.89)