Modeling the effects of cervical cancer prevention in the Netherlands

Vink, M.A.

2015

document version
Publisher's PDF, also known as Version of record

5 | Joint analysis of the natural antibody response to human papillomavirus (HPV) types 16 and 18 in a cross-sectional serological survey amongst Dutch women

MA Vink[1,2]
J Berkhof[2]
M van Boven[1]
JA Bogaards[1]

[1] Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands
[2] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, the Netherlands

Submitted

ABSTRACT

Background

Serological surveys have proved useful for studying infectious disease dynamics, and may be employed to assess the population effects of vaccination against human papillomavirus (HPV) types 16 and 18. We propose a joint analysis of HPV16 and HPV18 antibody concentrations to obtain insight into the association in natural antibody responses and to improve estimation of co-occurrence of both HPV types.

Methods

We analyzed concentrations of IgG HPV16 and HPV18 antibodies in serum, as measured by a VLP-based multiplex immunoassay, from a representative sample of the Dutch female population prior to introduction of HPV vaccination. We developed a four-component bivariate mixture model (representing individuals seropositive for none, one, or both HPV types) with age-dependent mixing proportions and possibly correlated antibody concentrations per mixture component. Various model assumptions were tested to assess associations in seropositivity and antibody responses between HPV16 and HPV18.

Results

The data were best described by a model with strongly correlated antibody concentrations in the double seronegative and double seropositive components. Classification of individuals seropositive for either one or both HPV types was improved in the bivariate model compared with a combination of univariate models per HPV type. HPV16 and HPV18 seropositivity was strongly associated, but the odds ratio declined from 64 for 10-20 year-olds to between 4 and 6 for older age groups.

Conclusions

A joint analysis of HPV16 and HPV18 natural antibody concentrations provides more reliable classification of single and double seropositive individuals than separate univariate analyses. Seroepidemiologic studies can be used to study co-occurrence of HPV types and may aid in interpreting the population effects of HPV vaccination from post-vaccine serological surveys.

INTRODUCTION

Human papillomavirus (HPV) is one of the most prevalent sexually transmitted infections. Persistent infection with high-risk HPV is the necessary cause for the development of cervical cancer,[1] which was ranked 4th on the worldwide cancer incidence in women in 2012.[2] HPV16 and HPV18 are responsible for approximately 70% of all cervical cancer cases. Reducing the cervical cancer incidence has been the main objective for the introduction of HPV16/18 vaccination in many developed countries. The Netherlands added HPV16/18 vaccination of 12-year-old girls to the Dutch national immunization program in 2009. The effects of vaccination on the cervical cancer incidence are not expected within the coming decades because the development of cervical cancer after an HPV infection takes over 25 years on average.[3] Monitoring the effectiveness of vaccination by surrogate endpoints is required, for example by means of testing for antibodies to viral antigens in serum, indicating past infection or vaccination.

Serological surveys are relatively inexpensive and only a small amount of serum is necessary to test for antibodies against a variety of pathogens. These surveys can be used for monitoring the antibody levels in vaccinated individuals, and also inform on the post-vaccine change in infection dynamics in the unvaccinated population, the so-called herd effect of mass immunization. This makes serology an interesting means for monitoring HPV vaccination, which is currently administered to Dutch girls only and for which vaccine uptake is relatively low at around 60% in the Netherlands.[4] Two population-based serological surveys have been carried out prior to the introduction of HPV vaccination in the Netherlands, and a third one is scheduled for 2016/2017.

Use of serology for monitoring HPV vaccination is complicated by two facts: the antibody response to infection with HPV is generally weak; and seropositivity is not considered a correlate of protection. Currently used serological assays are capable of detecting vaccine-induced antibodies, but the results regarding natural antibody levels are difficult to interpret due to a poor signal to noise ratio.[5] Classification of individuals as either seropositive or seronegative by means of a cut-off value leads to substantial misclassification bias, as the distribution of natural antibody concentrations of both groups strongly overlap.[6] A probabilistic assignment to either group on the basis of a mixture model improves estimation of seroprevalence as a function of age, but extending this methodology to multiple HPV types in not straightforward. To this end, we propose a bivariate mixture model, which jointly analyzes the antibody concentrations specific for two HPV types. Ideally, a joint analysis allows for estimating the association in natural antibody responses and in the occurrence of two HPV types while taking account of correlations induced by assay noise. The proposed methodology will further aid in the interpretation of HPV antibody detection, and is here applied to HPV16 and HPV18 specifically in light of monitoring the population effects of HPV16/18 vaccination by serological surveys.

METHODS

Serological data

We analyzed the IgG antibody concentrations to HPV16 and HPV18 collected in a large cross-sectional survey, representative of the Dutch general population. The samples were collected in 2006/2007, i.e. prior to the introduction of HPV16/18 vaccination in the Dutch national immunization program. A total of 3,875 randomly sampled women between 0 and 79 years of age provided a serum sample. HPV type-specific IgG antibodies against L1 virus-like-particles (VLP) were tested with a VLP-based multiplex immunoassay.[7] The assay measures the antibody concentrations to 7 high-risk HPV types simultaneously and it has a lower limit of detection for HPV16 at 0.08 luminex units per milliliter (LU/ml) and at 0.03 LU/ml for HPV18.

Previous analysis by a univariate mixture model

We have previously estimated HPV16 seroprevalence figures by fitting a univariate mixture model to log-transformed HPV16 antibodies from the serological survey.[6] In short, this mixture model consisted of two normally distributed mixture components, representing persons seronegative or seropositive for HPV16. Age-dependent mixing proportions, denoting HPV seroprevalence, were estimated simultaneously with the parameters of the component densities. We repeated the univariate analyses for HPV18 antibodies and present the seroprevalence figures (both smoothed and discretized by age group) in the supplementary material (Appendix Figure 1).

Bivariate mixture model

The log-transformed concentrations of HPV16 and HPV18 antibodies are jointly described by a mixture model with four bivariate normal component densities with unknown mean and covariance matrix. Each individual i with log-transformed HPV16 and HPV18 concentrations $y_i = (y_i^{16}, y_i^{18})$ contributes to the likelihood as follows:

$$f(y_i; p, \mu, \Sigma) = \sum_{k=1}^{4} p_k \phi(y_i; \mu_k, \Sigma_k),$$

with p_k denoting the mixing proportions. The bivariate normal component densities $\Phi(y_i; \mu_k, \Sigma_k)$ represent individuals that are:
Seronegative for both HPV types, with mean and covariance matrix:

$$\mu^{--} = \begin{pmatrix} \mu_{16}^{--} \\ \mu_{18}^{--} \end{pmatrix}, \ \Sigma^{--} = \begin{pmatrix} (\sigma_{16}^{--})^2 & \rho^{--}\sigma_{16}^{--}\sigma_{18}^{--} \\ \rho^{--}\sigma_{16}^{--}\sigma_{18}^{--} & (\sigma_{18}^{--})^2 \end{pmatrix}$$

Seropositive for HPV16 and seronegative for HPV18, with mean and covariance matrix:

$$\mu^{+-} = \begin{pmatrix} \mu_{16}^{+-} \\ \mu_{18}^{+-} \end{pmatrix}, \ \Sigma^{+-} = \begin{pmatrix} (\sigma_{16}^{+-})^2 & \rho^{+-}\sigma_{16}^{+-}\sigma_{18}^{+-} \\ \rho^{+-}\sigma_{16}^{+-}\sigma_{18}^{+-} & (\sigma_{18}^{+-})^2 \end{pmatrix}$$

Seropositive for HPV18 and seronegative for HPV16, with mean and covariance matrix:

$$\mu^{-+} = \begin{pmatrix} \mu_{16}^{-+} \\ \mu_{18}^{-+} \end{pmatrix},\ \Sigma^{-+} = \begin{pmatrix} (\sigma_{16}^{-+})^2 & \rho^{-+}\sigma_{16}^{-+}\sigma_{18}^{-+} \\ \rho^{-+}\sigma_{16}^{-+}\sigma_{18}^{-+} & (\sigma_{18}^{-+})^2 \end{pmatrix}$$

Seropositive for both HPV types, with mean and covariance matrix:

$$\mu^{++} = \begin{pmatrix} \mu_{16}^{++} \\ \mu_{18}^{++} \end{pmatrix},\ \Sigma^{++} = \begin{pmatrix} (\sigma_{16}^{++})^2 & \rho^{++}\sigma_{16}^{++}\sigma_{18}^{++} \\ \rho^{++}\sigma_{16}^{++}\sigma_{18}^{++} & (\sigma_{18}^{++})^2 \end{pmatrix}$$

The superscripts of the means, standard deviations, and correlations (parameters μ, σ *and* ρ respectively) indicate to which mixture component the parameter belongs (e.g., the HPV16 seropositive and HPV18 seronegative mixture component is represented by: + -), and the subscript stands for the HPV type.

The mixing proportions p_κ were taken to be age-dependent with $0 < p_\kappa(a) < 1$, $\sum_{\kappa=1}^{4} p_\kappa(a) = 1$, and divided into 5 age groups (0-10, 10-20, 20-40, 40-60, and 60-80 year olds). Because a survey on sexual health of youth in the Netherlands showed that sexual intercourse below age 12 is rare,[8] we assumed the antibody concentrations of 0–10-year-olds to inform only the double seronegative component density (i.e., $p_1 = 1$, $p_2 = p_3 = p_4 = 0$ for this age class).

Parameter estimation

Parameters were estimated in a Bayesian framework through Gibbs sampling using Just Another Gibbs Sampler (JAGS).[9] To avoid label switching, the positive mixture means were reparameterized:

$$\begin{pmatrix} \mu_{16}^{+-} \\ \mu_{18}^{+-} \end{pmatrix} = \begin{pmatrix} \mu_{16}^{--} + \Delta_{16}^{+-} \\ \mu_{18}^{+-} \end{pmatrix},$$

$$\begin{pmatrix} \mu_{16}^{-+} \\ \mu_{18}^{-+} \end{pmatrix} = \begin{pmatrix} \mu_{16}^{-+} \\ \mu_{18}^{--} + \Delta_{18}^{-+} \end{pmatrix},$$

$$\begin{pmatrix} \mu_{16}^{++} \\ \mu_{18}^{++} \end{pmatrix} = \begin{pmatrix} \mu_{16}^{--} + \Delta_{16}^{++} \\ \mu_{18}^{--} + \Delta_{18}^{++} \end{pmatrix},$$

with $\Delta_{16}^{+-}, \Delta_{18}^{-+}, \Delta_{16}^{++}, \Delta_{18}^{++} > 0$.

We took normal prior distributions for the negative mixture means (μ^{--}) and half-normal priors for the Δ's. Uniform prior distributions were taken for the standard deviations and the correlation parameters, and a Dirichlet prior was assumed for the mixing proportions. We ran four parallel Markov chain Monte Carlo (MCMC) models. For each chain, 500 iterations were taken as burn in and 12,500 iterations for sampling from the posterior distributions. We retained every 10th iteration, yielding a sample of 5,000 iterations. Convergence of the MCMC chains was inspected visually.

Imputation of censored concentrations

For the bivariate analysis we imputed antibody concentrations that fell below the detection limit of the serological assay, because JAGS could not handle partly missing observations. Twenty

women (0.5%) had censored concentrations for both HPV types, 156 women (4%) had a censored HPV16 concentration, and 32 women (0.8%) had a censored HPV18 concentration. For imputing the censored observations, we used the means and standard deviations from the HPV type-specific univariate 2-component mixture models as mentioned before.

If women had two censored antibody concentrations, we randomly drew an HPV16 and HPV18 log-transformed antibody concentration from the tail of a bivariate normal distribution (below the detection limit of the assay), using the pmvnorm function from the mvtnorm package in the statistical software R, assuming:

$$\mu = \begin{pmatrix} \mu_{16^-} \\ \mu_{18^-} \end{pmatrix},\ \Sigma = \begin{pmatrix} \sigma^2_{16^-} & \rho\sigma_{16^-}\sigma_{18^-} \\ \rho\sigma_{16^-}\sigma_{18^-} & \sigma^2_{18^-} \end{pmatrix}.$$

Here, μ_{16^-},μ_{18^-},σ_{16^-} and σ_{16^-} were taken to be the estimated mean (μ) and standard deviation (σ) of the seronegative component density from the univariate 2-component mixture model, and ρ is the correlation in the data excluding women with a censored observation.

If the antibody concentration for HPV16 was censored, we randomly drew a point below the HPV16 detection limit from the conditional normal distribution. In case the person's HPV18 concentration was more likely to come from the seronegative component than from the seropositive component, the conditional mean and standard deviation were taken to be:

$$\mu = \mu_{16^-} + \frac{\sigma_{16^-}}{\sigma_{18^-}}\rho(y_i^{18} - \mu_{18^-}),$$
$$\sigma^2 = (1-\rho^2)\sigma^2_{16^-}.$$

If the person's HPV18 concentration was more likely to come from the seropositive component density we took the conditional mean and standard deviation:

$$\mu = \mu_{16^-} + \frac{\sigma_{16^-}}{\sigma_{18^+}}\rho(y_i^{18} - \mu_{18^+}),$$
$$\sigma^2 = (1-\rho^2)\sigma^2_{16^-}.$$

Censored HPV18 concentrations were imputed analogously.

Model scenarios

We evaluated several mixture models to describe the HPV16 and HPV18 antibody concentrations. For all scenarios we assume type-specific antibody concentrations among seronegative individuals to be invariant to the serostatus of the other HPV-type i.e., $\mu^{--}_{16} = \mu^{-+}_{16}$, $\mu^{--}_{18} = \mu^{+-}_{18}$, $\sigma^{--}_{16} = \sigma^{-+}_{16}$ and $\sigma^{--}_{18} = \sigma^{+-}_{18}$. This can be interpreted as an absence of mutual cross-reactivity between natural antibodies induced by either HPV16 or HPV18 infection.

We evaluated five bivariate mixture models that differ with respect to the assumptions of the bivariate normal densities and the mixture weights. In Table 1 we present an overview of the assumptions of each of the model scenarios: in Scenario 1 we assumed the HPV16 and HPV18 antibody concentrations to be independent, that there was no association in the occurrence

of HPV types, and that each HPV type would be described by a marginal distribution with two mixture components. The Scenario 1 model is basically the product of two univariate models, one for HPV16 and one for HPV18.[6] We relaxed the assumption of independent occurrence in Scenario 2. In Scenario 3 we allowed for linear dependence in antibody concentrations per mixture component. Scenario 4 allowed for dependence in antibody concentrations and in occurrence, but we retained the constraint of a marginal description by two components for each HPV type. We relaxed this constraint in Scenario 5: the seropositive means may be different for the double positive and single positive mixture components, e.g. due to boosting of antibody concentrations upon multiple infections.

Table 1. Overview of the model assumptions of five bivariate mixture models for describing the HPV16 and HPV18 antibody concentrations.

	Correlation between HPV16 and HPV18 antibody concentrations	Association in occurrence of HPV16 and HPV18	Marginal number of mixture components per HPV type	Total number of parameters to be estimated
Scenario 1	No correlation; $p_\kappa = 0$ for all k	Independence in occurrence of HPV16 and HPV18; p+-=p+· p-· with p+·/ p-· the marginal proportions	Two marginal components; $\mu_{16}^{++} = \mu_{16}^{+-}, \mu_{18}^{++} = \mu_{18}^{-+}$, $\sigma_{16}^{++} = \sigma_{16}^{+-}$, and $\sigma_{18}^{++} = \sigma_{18}^{-+}$	16
Scenario 2	No correlation; $p_\kappa = 0$ for all k	No restrictions, occurrence of HPV16 and HPV18 may be associated	Two marginal components	20
Scenario 3	No restrictions, antibody concentrations may be correlated.	Independence in occurrence of HPV16 and HPV18; p+-=p+· p-· with p+·/ p-· the marginal proportions	Two marginal components	20
Scenario 4	No restrictions, antibody concentrations may be correlated.	No restrictions, occurrence of HPV16 and HPV18 may be associated	Two marginal components	24
Scenario 5	No restrictions, antibody concentrations may be correlated.	No restrictions, occurrence of HPV16 and HPV18 may be associated	Three marginal components are allowed	28

We used the deviance information criterion (DIC) for model selection. The DIC balances the deviance and the number of parameters. The scenario with the lowest DIC is to be preferred on statistical grounds. Although there is not a formal threshold to assign a relevant difference between two models, a difference of more than 7 to 10 points is generally taken to favor the model with the smallest DIC.[10, 11]

We assessed whether the bivariate model (i.e., the scenario with the lowest DIC) would be better able to classify individuals into one of the four mixture components when compared with two univariate models for HPV16 and HPV18. We simulated 50 bivariate data sets of 3800

individuals (corresponding to the size of the serological data set) using the parameter estimates of the selected bivariate model. Each simulated data set was fitted by both the bivariate model and by a univariate two-component mixture model for each HPV type. Per data set, we calculated for each individual the probability of belonging to one of the four mixture components by Monte Carlo integration.[12] We assigned an individual to the component that had largest support and compared how many individuals were correctly classified by the bivariate model and by the combination of two univariate models.

RESULTS

We observed a strong correlation between the log-transformed HPV16 and HPV18 responses across the range of antibody concentrations (Figure 1). These concentrations increased with age (Appendix Figure 2). Model scenarios that allowed for correlation between the log-transformed HPV16 and HPV18 antibody responses (Scenarios 3-5) invariably outperformed the models that assumed no correlation (Scenarios 1-2), as judged by the DIC (Table 2). Within the set of scenarios that accounts for the association in HPV16 and HPV18 seropositivity (Scenarios 2, 4, 5), describing the single seropositive and the double seropositive component densities by different positive mixture means (Scenario 5) gave the best description of the serological data. We present parameter estimates for Scenario 5 together with their 95% credible intervals in Appendix Table 1, and the model fit by contour lines of the mixture density to a heat plot of the data in Figure 1.

HPV16 and HPV18 antibody concentrations were strongly correlated, with Pearson correlation coefficients of 0.7 for both the double seronegative and double seropositive component densities. The correlations in the single seropositive components were substantially lower ($\rho^{+-} = 0.4$

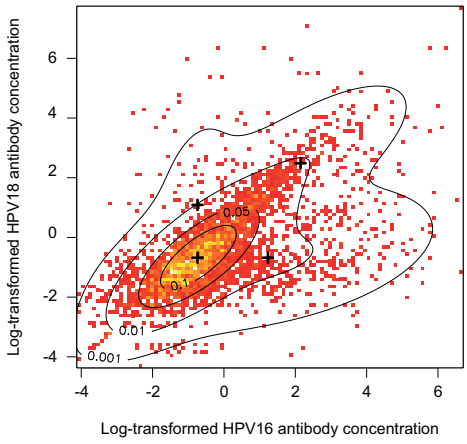


Figure 1. Heat plot of the log-transformed antibody concentrations to HPV16 and HPV18 together with the contour lines of the preferred bivariate mixture model (parameter medians of Scenario 5). The crosses denote the location of the mixture component means.

Table 2. Median parameter values of the bivariate normal component densities and deviance information criterion (DIC) per evaluated scenario.

	HPV16- HPV18-	HPV16+ HPV18-	HPV16- HPV18+	HPV16+ HPV18+	DIC
Scenario 1	$N\left(\begin{pmatrix}-0.6\\-0.71\end{pmatrix}\begin{pmatrix}1.15^2 & 0\\0 & 1.09^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}2.49\\-0.71\end{pmatrix}\begin{pmatrix}1.46^2 & 0\\0 & 1.09^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}-0.66\\2.12\end{pmatrix}\begin{pmatrix}1.15^2 & 0\\0 & 1.58^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}2.49\\2.12\end{pmatrix}\begin{pmatrix}1.46^2 & 0\\0 & 1.58^2\end{pmatrix}\right)$	41068.1
Scenario 2	$N\left(\begin{pmatrix}-0.77\\-0.80\end{pmatrix}\begin{pmatrix}1.18^2 & 0\\0 & 1.06^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.99\\-0.80\end{pmatrix}\begin{pmatrix}2.37^2 & 0\\0 & 1.06^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}-0.77\\1.53\end{pmatrix}\begin{pmatrix}1.18^2 & 0\\0 & 2.65^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.99\\1.53\end{pmatrix}\begin{pmatrix}2.37^2 & 0\\0 & 2.65^2\end{pmatrix}\right)$	39987.0
Scenario 3	$N\left(\begin{pmatrix}-0.65\\-0.57\end{pmatrix}\begin{pmatrix}1.16^2 & 1.12\\1.12 & 1.21^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.09\\-0.57\end{pmatrix}\begin{pmatrix}2.10^2 & 1.11\\1.11 & 1.21^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}-0.65\\1.34\end{pmatrix}\begin{pmatrix}1.16^2 & 0.96\\0.96 & 2.49^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.09\\1.34\end{pmatrix}\begin{pmatrix}2.10^2 & 4.52\\4.52 & 2.49^2\end{pmatrix}\right)$	37623.7
Scenario 4	$N\left(\begin{pmatrix}-0.75\\-0.69\end{pmatrix}\begin{pmatrix}1.10^2 & 0.90\\0.90 & 1.11^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.62\\-0.69\end{pmatrix}\begin{pmatrix}1.87^2 & 0.91\\0.91 & 1.11^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}-0.75\\1.97\end{pmatrix}\begin{pmatrix}1.10^2 & 0.69\\0.69 & 1.74^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.62\\1.97\end{pmatrix}\begin{pmatrix}1.10^2 & 2.86\\2.86 & 1.74^2\end{pmatrix}\right)$	37347.8
Scenario 5	$N\left(\begin{pmatrix}-0.74\\-0.68\end{pmatrix}\begin{pmatrix}1.09^2 & 0.89\\0.89 & 1.10^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}1.23\\-0.68\end{pmatrix}\begin{pmatrix}2.07^2 & 0.90\\0.90 & 1.10^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}-0.74\\1.09\end{pmatrix}\begin{pmatrix}1.09^2 & 0.82\\0.82 & 2.59^2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}2.15\\2.49\end{pmatrix}\begin{pmatrix}1.42^2 & 1.28\\1.28 & 1.24^2\end{pmatrix}\right)$	37311.7

Scenario 1: assumed independence between HPV16/18 seroprevalence and between HPV16/18 antibody concentrations, and two locations of the component means per HPV type

Scenario 2: assumed independence between HPV16/18 antibody concentrations, and two locations of the mixture means per HPV type

Scenario 3: assumed independence between HPV16/18 seroprevalence, and two locations of the mixture means per HPV type

Scenario 4: assumed two locations of the component means per HPV type

Scenario 5: assumed three locations of the component means per HPV type

and $\rho^{+} = 0.3$). The age-dependent mixing proportions, representing the population prevalence per component, are presented in Figure 2. Single seropositivity occurred more often for HPV16 than for HPV18; on average 31% of the HPV16 seropositive persons were seropositive for HPV18, while on average 71% of the HPV18 seropositive persons also were HPV16 seropositive. There is a strong association in seropositivity for HPV types 16 and 18. This is confirmed by the odds ratio (OR) that we obtained from the mixing proportions; we considered these as age-dependent 2-by-2 contingency tables on an aggregate level. The association of HPV16 and HPV18 seropositivity was especially large for the age group 10-20 years (OR = 64), while it declined for older age groups: OR = 6.4 for 20-40 year-olds, OR = 5.4 for 40-60 year-olds, and OR = 4.2 for the oldest age group (see Table 3). Note that the association in HPV16 and HPV18 seropositivity was significant in all age groups.

The estimated marginal HPV16 seroprevalence in the bivariate analysis was substantially higher than in the univariate HPV16 analysis (Figure 3). Specifically, the difference in seroprevalence was largest for age group 40-60 years with an absolute difference of 13.5% in point estimates. Notice that the marginal fit to the data was very similar for the univariate and bivariate models (Appendix Figure 3). We assessed whether the bivariate model was better able to classify individuals in the correct mixture component compared with a combination of univariate analyses for HPV16 and HPV18, and found a better classification by the bivariate model in 49/50 simulated data sets. The two models did equally well in classifying the double seronegative and double seropositive individuals: respectively 80% and 85% of individuals were correctly classified. However, the bi-

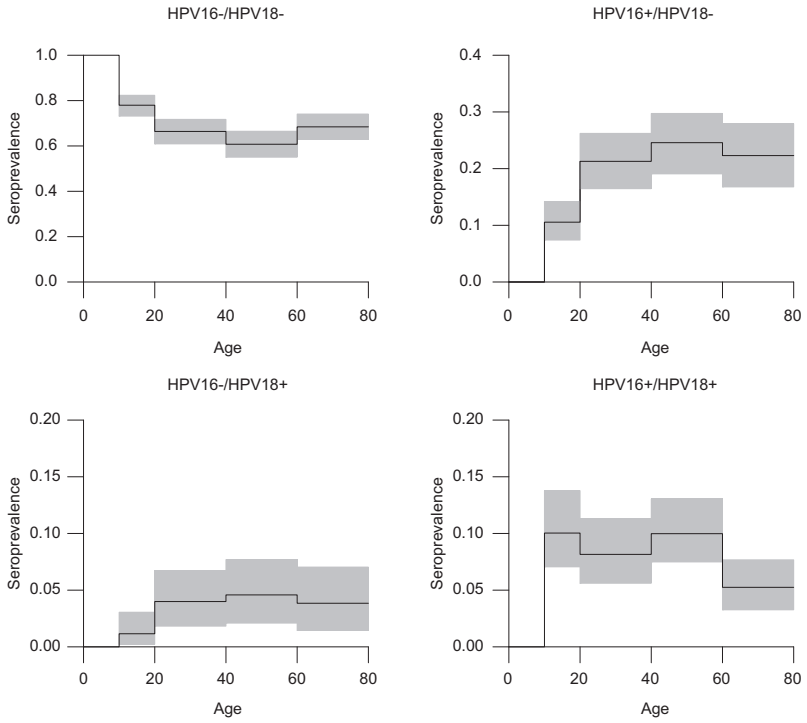


Figure 2. Estimated age-specific seroprevalence figures per mixture component for Scenario 5 (i.e., the model that allows for dependence between HPV16/18 antibody concentrations, dependence in occurrence, and for three marginal mixture components). The solid lines are the median values, the shaded area represent the 95% credible interval.

Table 3. Age-specific odds ratio (95% credible interval) of the association of HPV16 and HPV18 seropositivity.

Age group (years)	0-10	10-20	20-40	40-60	60-80
Odds ratio	NA	64.3 (21.4-355.5)	6.4 (3.2-15.0)	5.4 (2.8-12.8)	4.2 (1.8-12.0)

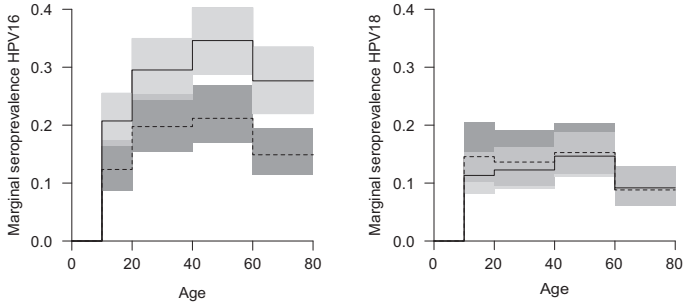


Figure 3. Marginal HPV16 and HPV18 seroprevalence in women. The dashed lines denote the seroprevalence from the univariate mixture model, the solid lines are from the preferred bivariate mixture model (Scenario 5). The lines are the median values, the shaded area represent the 95% credible interval.

variate model provided a substantially better classification of the single seropositives: 64% versus 52% for single HPV16 seropositive individuals and 59% versus 51% for single HPV18 seropositives.

DISCUSSION

In this paper, we propose a bivariate mixture model for joint analysis of antibody levels against distinct genotypes in cross-sectional serological surveys. The method is applied to HPV16 and HPV18 antibody concentrations in women collected prior to the introduction of HPV16/18 vaccination in the Netherlands. Our analysis demonstrated strongly correlated antibody concentrations for double seronegative and double seropositive individuals, but less so in single seropositive individuals. In addition, double seropositive individuals had higher antibody concentrations than single seropositive individuals. We estimated a strong association between HPV16 and HPV18 seropositivity, especially amongst teenagers, which is significant across all age groups.

Our method provides a full description of the HPV16/18 antibody concentrations and allows for post-hoc assessment of the age-specific association in HVP16 and HPV18 seropositivity. Age-specific associations in co-occurrence have mostly been dealt with using methods that rely on multivariate binary data (i.e., denoting seronegative or seropositive for each type separately). These methods model the odds ratio of co-occurrence either by means of a parametric link function as in the bivariate Dale model,[13, 14] or via frailty models.[15, 16] However, in case of natural antibodies against HPV, seropositivity needs to be inferred from a mixture model. The use of a cut-off value may lead to a substantial misclassification bias for HPV and would consequently lead to biased estimates of the age-specific odds of co-occurrence. In addition, the choice of a parametric link function or of the frailty distribution may be complicated if the odds ratios as function of age are very peaked, as we show to be the case for HPV16 and HPV18 seropositivity. Our analysis circumvents this problem as we model the age-specific mixing proportions independent of each other conditional on the mixture densities, which gives an unconstrained estimate of co-occurrence.

Our analysis was stratified into age groups of 10 to 20 years to account for changes in seroprevalence and co-occurrence by age. We showed for the univariate analyses that the stratified seroprevalence resembled the smoothed seroprevalence figures. An obvious extension of the bivariate model would involve modeling the mixing proportions by a flexible but age-dependent function, for example by means of Penalized splines as in our univariate mixture model published before.[6] This would be especially worthwhile for applications to smaller data sets or with an uneven distribution of serum samples across age groups. Such an extension would also yield a better resolution with respect to the age-specific pattern of co-occurrence between HPV16 and HPV18 seropositivity.

The age-specific odds-ratios reflect the heterogeneity in sexual behavior associated with the transmission of HPV16 and HPV18. Clustering of HPV types based on seropositivity status is expected as persons who have been infected with one type are at increased risk for acquisition

of a second type[17] and this has indeed been confirmed by serological data analyses.[18-21] To our knowledge, we are the first to present the age-specific association in occurrence of HPV16 and HPV18 seropositivity. Our results suggest that clustering of HPV16 and HPV18 infection is particularly large among teenagers. This could indicate that those teenagers who become sexually active at an early age have a high chance of becoming infected with both HPV16 and HPV18, e.g. due to frequent partner change. The substantially smaller odds ratios at older ages show that the population becomes more homogeneous with respect to sexual history.

Our analyses yielded estimates of the correlation of antibody concentrations in the four mixture components. The strong correlation among double seronegative individuals presumably represents assay noise, given that seronegativity is related to absence of prior exposure to HPV16/18, certainly at young age. As we expect that double seropositive persons have experienced both HPV16 and HPV18 infection, the correlation in the double seropositive component could also reflect the immunologic profile of a host's response to infection; a strong antibody response to HPV16 could imply a strong response to HPV18. Additionally, we found that the antibody concentrations of the double seropositive individuals were larger than those measured for the same HPV type in single seropositive individuals. Possibly, the assay targets not only antibodies to the main L1 epitope but also recognizes epitopes that are shared between HPV16 and HPV18, which results in a boosted immune response to one type after infection with the other type.[22]

The marginal HPV16 seroprevalence was substantially larger in the bivariate analysis than in the univariate analysis, even though the two model fits to the marginal data were very much alike. Both models seem to describe the data equally well, but from the simulation study we concluded that the bivariate mixture model is better able to classify individuals in the correct quadrant. The bivariate model enables more flexibility in describing seropositive antibody concentrations by the introduction of a third mixture component, stratified by HPV18 serostatus. Extending the number of mixture components beyond two in a univariate analysis would not have straightforward interpretation. By accounting for more heterogeneity in describing the off-diagonal antibody concentrations, the bivariate model could better classify the single seropositive individuals that had lower antibody concentrations compared with the double seropositive persons. Indeed, the increased HPV16 seroprevalence in the bivariate model was predominantly due to reclassification of samples with HPV16-specific antibody concentrations that had similar support for seronegativity or seropositivity in the univariate analysis. It should be investigated whether this difference in marginal seroprevalence is problematic for monitoring studies that compare pre- and post-vaccine seroprevalence figures to assess the effects of HPV vaccination.

Introduction of HPV16/18 vaccination in the national immunization program will lead to less circulation of vaccine types through the population. Consequently, the infection risk in the unvaccinated population will decrease resulting in a reduced risk for the development of cervical cancer. A shift in serostatus, from seropositive to seronegative, on the population level is expected for each type, and this has to be monitored for the partly vaccinated cohorts. Estimated trends in seroprevalence from post-vaccine surveys may depend on the method that is used to analyze the serological data. A joint analysis enables accurate assessment of the proportion of

individuals that is seropositive for one or both HPV types. Moreover, our model generates insight into correlated detection of antibodies with multiplex immunoassay technology. This study shows that HPV serology can be used for epidemiologic studies with an interest in co-occurrence of HPV types and may aid in interpreting post-vaccine serological surveys for assessing the overall impact of HPV vaccination.

ACKNOWLEDGEMENTS

We thank Dr. J. van de Kassteele for his help with programming in JAGS.

REFERENCES

1. Nobbenhuis MA, Walboomers JM, Helmerhorst TJ, et al. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. *Lancet.* 1999;354(9172): 20-25.
2. Cancer Statistics UK. Cervical Cancer Key Facts. Retrieved 24 October 2014, from www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/cervical-cancer/.
3. Vink MA, Bogaards JA, van Kemenade FJ, et al. Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data. *Am J Epidemiol.* 2013;178(7): 1161-1169.
4. van Lier EA, Oomen PJ, Mulder M, et al. Immunisation coverage National Immunisation Programme in the Netherlands: Year of report 2013, National Institute for Public Health and the Environment. 2013.
5. Stanley M. Immunobiology of HPV and HPV vaccines. *Gynecol Oncol.* 2008;109(2 Suppl): S15-21.
6. Vink MA, van de Kassteele J, Wallinga J, et al. Estimating Seroprevalence of Human Papillomavirus Type 16 Using a Mixture Model with Smoothed Age-dependent Mixing Proportions. *Epidemiology.* 2015;26(1): 8-16.
7. Scherpenisse M, Mollers M, Schepp RM, et al. Seroprevalence of seven high-risk HPV types in The Netherlands. *Vaccine.* 2012;30(47): 6686-6693.
8. de Graaf H, Meijer S, Poelman J, et al. Sex Below the Age of 25: Sexual Health of Youth in the Netherlands in 2005 [in Dutch]. Delft, the Netherlands, Eburon Academic Publishers2005.
9. Plummer M. JAGS: A program for analysis of Bayesian graphical models usings Gibbs sampling. 2003.
10. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model comlexity and fit. *J R Stat Soc Ser B (Statistical Methodol).* 2002;64(4): 583-639.
11. Lesaffre E (2012). Model building and assessment. Bayesian biostatistics. E. Lesaffre and A. B. Lawson. New York, John Wiley & Sons**:** 267-318.
12. Gelfand AE and Smith AFM. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc.* 1990;85(410): 398-409.
13. Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics.* 1986;42(4): 909-917.
14. Hens N, Aerts M, Shkedy Z, et al. Modelling multisera data: the estimation of new joint and conditional epidemiological parameters. *Stat Med.* 2008;27(14): 2651-2664.
15. Unkel S and Farrington CP. A new measure of time-varying association for shared frailty models with bivariate current status data. *Biostatistics.* 2012;13(4): 665-679.
16. Farrington CP, Whitaker HJ, Unkel S, et al. Correlated infections: quantifying individual heterogeneity in the spread of infectious diseases. *Am J Epidemiol.* 2013;177(5): 474-486.
17. Vaccarella S, Franceschi S, Snijders PJ, et al. Concurrent infection with multiple human papillomavirus types: pooled analysis of the IARC HPV Prevalence Surveys. *Cancer Epidemiol Biomarkers Prev.* 2010;19(2): 503-510.
18. Vaccarella S, Franceschi S, Clifford GM, et al. Seroprevalence of antibodies against human papillomavirus (HPV) types 16 and 18 in four continents: the International Agency for Research on Cancer HPV Prevalence Surveys. *Cancer Epidemiol Biomarkers Prev.* 2010;19(9): 2379-2388.
19. Vriend HJ, Bogaards JA, van der Klis FR, et al. Patterns of human papillomavirus DNA and antibody positivity in young males and females, suggesting a site-specific natural course of infection. *PLoS One.* 2013;8(4): e60696.
20. Combes JD, Pawlita M, Waterboer T, et al. Antibodies against high-risk human papillomavirus proteins as markers for invasive cervical cancer. *Int J Cancer.* 2014;135(10): 2453-2461.

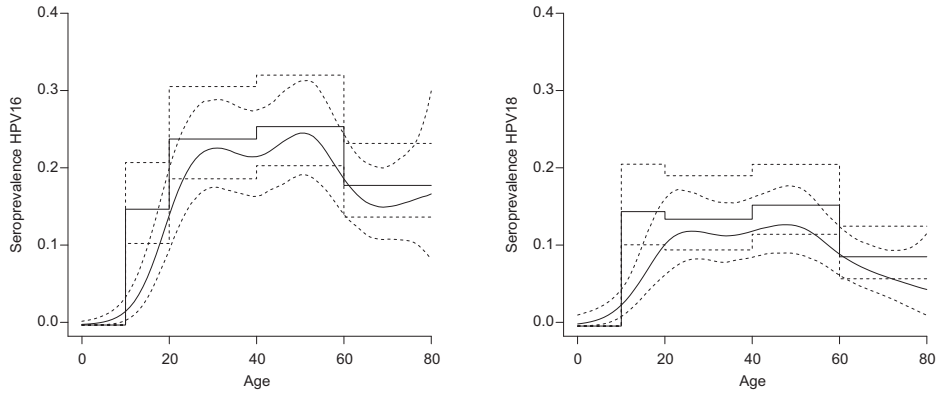
21. Syrjanen S, Waterboer T, Kero K, et al. Oral human papillomavirus infection in men might contribute to HPV serology. *Eur J Clin Microbiol Infect Dis.* 2014.

22. Scherpenisse M, Schepp RM, Mollers M, et al. Characteristics of HPV-specific antibody responses induced by infection and vaccination: cross-reactivity, neutralizing activity, avidity and IgG subclasses. *PLoS One.* 2013;8(9): e74797.

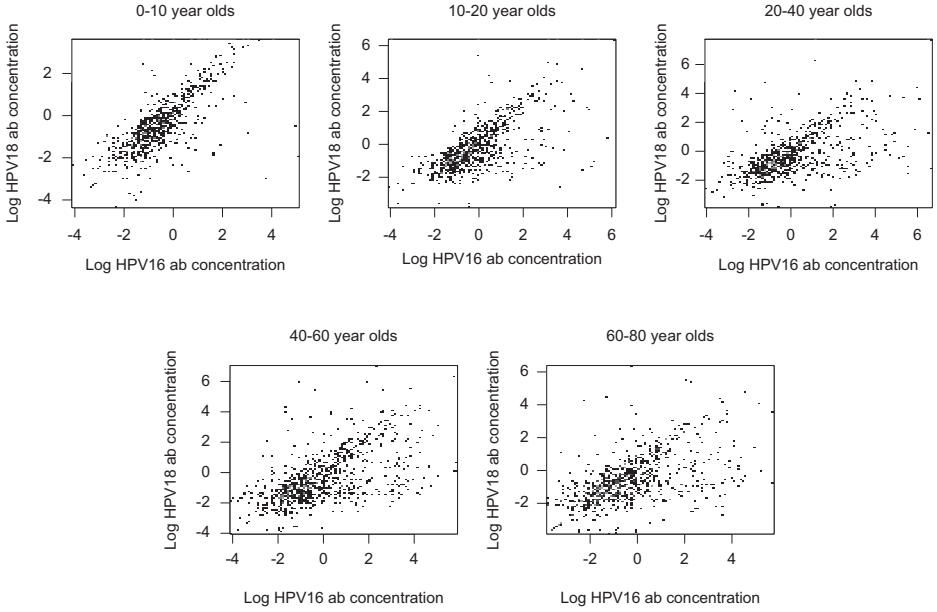
APPENDIX

Appendix Table 1. Estimated parameters for Scenario 5 (median and 95% credible interval).

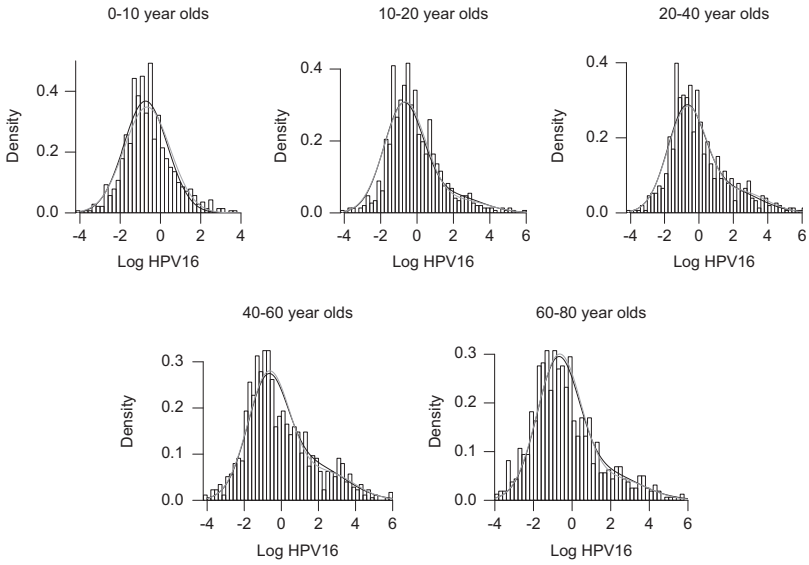
	HPV16- HPV18-		HPV16+ HPV18-		HPV16- HPV18+		HPV16+ HPV18+	
	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI
μ_{16}	-0.74	(-0.79, -0.7)	1.23	(0.93, 1.72)	-0.74	(-0.79, -0.7)	2.15	(1.78, 2.48)
μ_{18}	-0.68	(-0.72, -0.64)	-0.68	(-0.72, -0.64)	1.09	(0.28, 3.15)	2.49	(2.14, 2.78)
σ_{16}	1.09	(1.05, 1.21)	2.07	(1.86, 2.23)	1.09	(1.05, 1.21)	1.42	(1.25, 1.62)
σ_{18}	1.10	(1.06, 1.13)	1.10	(1.06, 1.13)	2.59	(1.70, 3.08)	1.24	(1.08, 1.45)
ρ	0.75	(0.73, 0.77)	0.39	(0.30, 0.48)	0.29	(0.07, 0.52)	0.73	(0.58, 0.83)



Appendix Figure 1. Marginal pre-vaccine seroprevalence of HPV16 (left panel) and HPV18 (right panel) in women from a univariate analysis of the serological data. The seroprevalence is either modeled by a P-spline line as published before[6] (smoothed line) or discretized in five age groups (step function).



Appendix Figure 2. Log-transformed HPV16 and HPV18 antibody concentrations by age group.



Appendix Figure 3. Comparison of the log transformed HPV16 antibody concentrations per age group of the serological data (histogram) with the estimated marginal density of the univariate mixture model (gray line) and the bivariate mixture model: Scenario 5 (black line). Note that gray and black lines are almost indistinguishable.