

Recurring infection with ecologically distinct human papillomavirus (HPV) types explains high prevalence and diversity

Sylvia Ranjeva^{*a}, Edward B Baskerville^a, Vanja Dukic^c, Luisa Villa^{d,1},
Eduardo Lazcano-Ponce^{e,1}, Anna Giuliano^b, Greg Dwyer^a, Sarah Cobey^a

^aUniversity of Chicago, Chicago, IL, USA

^bH. Lee Moffitt Cancer Center, Tampa, FL, USA

^cUniversity of Colorado Boulder, Boulder, CO, USA

^dUniversidade de Sao Paulo, Sao Paulo, Brazil

^eInstituto Nacional de Salud Publica, Cuernavaca, Mexico

^{*}Corresponding Author: slr@uchicago.edu

¹ L.L.V and E.L.P contributed equally to the work

Abstract

The high prevalence of human papillomavirus (HPV), the most common sexually transmitted infection, arises from the coexistence of over 200 genetically distinct types. Accurately predicting the impact of vaccines that target multiple types requires understanding the factors that determine HPV diversity. The diversity of many pathogens is driven by type-specific or “homologous” immunity, which promotes the spread of variants to which hosts have little immunity. To test for homologous immunity and to identify mechanisms determining HPV transmission, we fitted nonlinear mechanistic models to longitudinal data on genital infections in unvaccinated men. Our results provide no evidence for homologous immunity, instead showing that infection with one HPV type strongly increases the risk of infection with that type for years afterwards. For HPV16, the type responsible for most HPV-related cancers, an initial infection increases the one-year probability of reinfection by 20-fold, and the probability of reinfection remains 14-fold higher two years later. This increased risk occurs in both sexually active and celibate men, suggesting that it arises from auto-inoculation, episodic reactivation of latent virus, or both. Overall our results show that high HPV prevalence and diversity can be explained by a combination of a lack of homologous immunity, frequent reinfections, weak competition between types, and variation in type fitness between host subpopulations. Due to the high risk of reinfection, vaccinating boys that have not yet been exposed may be crucial to reduce prevalence, but our results suggest that there may also be large benefits from vaccinating previously infected individuals.

Introduction

Human papillomavirus (HPV), a major cause of genital warts and anogenital and oropharyngeal cancers [46, 32, 118], is the most common sexually transmitted infection [119]. While genital HPV infects approximately 40% of women and 45% of men in the USA, over 200 HPV types have been identified, and the prevalence of individual types never exceeds 10% [44]. The quadrivalent vaccine has been effective against the HPV types that cause the greatest burden of disease [29, 70, 115], and a recent nine-valent vaccine covers additional oncogenic types [52]. Predicting the effects of multivalent vaccines on HPV prevalence, however, is difficult without defining the factors that underlie HPV transmission and diversity, which are poorly understood.

Epidemiological theory has shown that the abundance of many pathogens depends on the dynamics of competition for susceptible hosts [66, 112, 18]. Pathogen strains compete by inducing adaptive immune responses that are specific to shared antigens, limiting the growth rates of antigenically similar competitors [18]. The accumulation of specific immunity in the host population to common strains decreases the transmission of those strains and promotes the spread of rarer antigenic variants, a phenomenon known as frequency-dependent selection. Strains that vary antigenically are often organized into groups that vary in dominance in space and time, yielding complex patterns of coexistence [43, 66]. Such immune-mediated competition can partly explain the antigenic and genetic diversity of influenza, pneumococcus, rotavirus, norovirus, *Neisseria meningitidis*, malaria, hepatitis C, HIV, trypanosomes, and other common pathogens [71, 18, 66, 65, 112].

It is unclear how HPV interacts with the immune system during infection, but in principle, distinctions between HPV types may arise due to acquired immunity from B cells and T cells. HPV types are defined by a 10% threshold of dissimilarity in the L1 gene, which codes for the major capsid protein [19]. The outer capsid modulates viral entry into host cells at the epithelial basement membrane [13], and the humoral response to infection is mainly type-specific anti-L1 antibodies [92, 108]. Studies of HPV in T-cell-deficient people show that cellular immunity is important in the control of infection [103, 92, 6, 93, 31], but it remains unclear how cellular immunity achieves type-specific recognition, if at all. In individuals with cancerous lesions from HPV16, cytotoxic T lymphocytes specific to the E6 and E7 oncoproteins are correlated with reduced disease [82, 95]. The specificity of the T cell repertoire to other genes or to the majority of HPV types, however, is not well established.

Efforts to understand the effects of immunity on HPV dynamics must begin with homologous immunity, or protection against repeat infection with the same HPV type. Homologous immunity would limit the prevalence of any type through negative frequency dependence. The traditional assumption is that most HPV-infected individuals permanently clear infection after 1–2 years [107, 100, 78], suggesting protective homologous immunity. In this scenario, the elevated cancer risk associated with particular HPV types results from a small fraction of persistent infections [78]. Longitudinal studies, however, have shown that hosts can be infected repeatedly [104, 28, 29]. Although there is some evidence that type-specific antibodies might provide modest protection against future infection in women [6, 28], serum antibody is not a marker of immune protection in men [90, 67]. Thus, the form and strength of homologous immunity remain unclear.

Evidence for competition between HPV types is also conflicting. Immune-mediated competition has been invoked to explain small increases in the prevalence of non-vaccine HPV types following HPV vaccination [53], and to explain weak cross-protection from the vaccine against related non-vaccine types [121]. Mathematical models show that type competition is consistent with observed patterns of HPV prevalence [25, 81]. Nevertheless, there is little empirical evidence for inter-type competition [64, 98], as shown in part by elevated rates of multiple-type compared to single-type infections and frequent concurrent acquisition of HPV types [15, 114]. Type competition may therefore be absent or weak [99, 16].

Meanwhile, the risk of HPV infection appears to depend on differences in demographic and behavioral risk factors between host subpopulations. For example, a host's number of sexual partners strongly affect infection risk [122, 33, 36, 79, 78], and some evidence suggests that oncogenic and non-oncogenic types are affected differently by variation in numbers of partners among hosts [60, 111]. However, detailed comparisons of risk factors for infection with each HPV type have not been performed. Different risk factors would suggest differences in transmission rates between types among host subpopulations, defining ecological distinctions that could potentially explain type prevalence.

To investigate the factors that determine HPV prevalence and diversity, we fitted mechanistic models of HPV dynamics to an extensive longitudinal dataset. Mechanistic models have long been used in infectious disease ecology to quantify the biological processes that underly pathogen dynamics, but HPV has received comparatively little attention [100]. Mechanistic models of HPV have generally focused on qualitative dynamics [76, 80, 12, 96, 89, 25, 57, 8, 4, 14, 117, 50, 101] and on predictions for health and economic policy [8, 105, 56, 24, 11, 116, 12], relying on informal methods to estimate parameters. Detailed longitudinal studies, however, present an opportunity to use mechanistic models to rigorously test hypotheses using robust statistical methods. Here we use such methods to show that the prevalence and diversity of HPV types are best explained by the combined effects of a lack of competition, whether within or between types, high rates of reinfection or persistence within individuals, and modest differences in high-risk subpopulations between types.

Results

Low-prevalence HPV types coexist over time

We fitted models to data from the HPV in Men (HIM) study, which tracks genital HPV infection and demographic and behavioral traits in unvaccinated men sampled at six month intervals over five years [34, 35, 33]. These data show that the mean prevalence of all HPV is 65% but that no type has a mean prevalence greater than 10% (Figure 1A), and that the prevalence of individual types is roughly constant over time (Figure 1B), demonstrating coexistence. The rank prevalence of HPV types is also constant among geographical locations, although absolute prevalence is higher in Brazil than in the USA or Mexico (Figure S3). We analyzed the five HPV types with the highest mean prevalences: HPV62, HPV84, HPV89, HPV16, and HPV51. We also analyzed HPV6, the ninth most prevalent type, because of its clinical significance. HPV6 is included in the quadrivalent vaccine and is one of two HPV types that are together associated with 90% of genital warts [45]. In our models, we accounted for risk factors that have been previously identified as being associated with the incidence and duration of HPV infection in the HIM dataset [34, 33, 86, 84, 94]. The risk factors for any HPV type include markers of increased exposure to infected sexual partners, which is not surprising for a sexually transmitted infection. For example, higher numbers of recent and lifetime sexual partners, either male or female, increase risk, whereas consistent condom use reduces risk among nonmonogamous men. In addition, risk factors include non-sexual behaviors, such as tobacco use, and demographic traits. For example, incidence is lower among ethnic Asians than other ethnic groups, and among individuals having completed education beyond secondary school. Several risk factors, notably circumcision and sexual orientation, differ in their effects among HPV types [1, 85]. Previous work thus highlighted the breadth of risk factors that affect HPV transmission, and type-specific differences in some of these risk factors would reflect meaningful ecological distinctions between HPV types. We therefore included the effects of a diverse set of risk factors in our models.

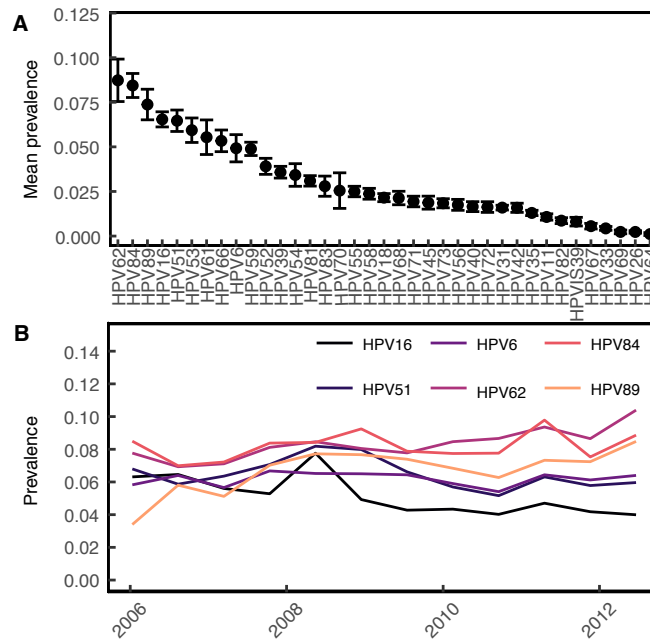


Figure 1: Prevalence of HPV types in the study population. **A** Average prevalence of each HPV type, in order of decreasing prevalence, with standard errors calculated across visits. **B** Prevalence over time of the six HPV types included in our analyses.

Past infection with HPV confers minimal protection against infection with the same type

Our models test three hypotheses about the dynamics of HPV infection. Under the simplest, or “memoryless” model, the risk of infection depends only on the effects of host- and HPV-type-specific risk factors (Equation 1), with no consideration of immunity. Two more complex models account for the effects of previous infection with the same type. In all models, we assumed that the parameters describing the distribution of the duration of carriage are fixed and type-specific. Our three models then differ only in their assumptions about the instantaneous per-capita infection risk, or “force of infection” [54]. The force of infection for host i with HPV type j at time t is given by λ_{ijt} :

- (i) In the memoryless model, risk of infection with type j depends only on host specific risk factors and the baseline force of infection for that type, λ_{0j} :

$$\lambda_{ijt} = \lambda_{0j} f(\vec{\alpha}_j^T \mathbf{C}_{it}), \quad (1)$$

The vector $\vec{\alpha}_j$ scales the effect of each of the M covariates on the force of infection, where \mathbf{C}_{it} is the covariate matrix. The effects of the covariates on the force of infection are represented by $f(\vec{\alpha}_j^T \mathbf{C}_{it})$, according to:

$$f(\vec{\alpha}_j^T \mathbf{C}_{it}) = e^{\alpha_{j,1} C_{1it} + \alpha_{j,2} C_{2it} + \alpha_{j,3} C_{3it} + \dots + \alpha_{j,M} C_{Mit}}. \quad (2)$$

- (ii) In the homologous immunity model, protective immunity reduces the probability of reinfection ($p_{\text{subsequent}}$) relative to the probability of an initial infection (p_{first}) with the same type:

$$\begin{aligned} \lambda_{ijt} &= \lambda_{0j} f(\vec{\alpha}_j^T \mathbf{C}_{it}), \\ p_{\text{first}} &= (1_{t+\Delta t} | 0_t) = (1 - e^{-\lambda_{ijt} \Delta t}), \\ p_{\text{subsequent}} &= p_{\text{first}} (1 - (1 - d) e^{-w(t - t_{\text{clr}})}), \end{aligned} \quad (3)$$

The strength of homologous immunity d is constrained to fall between 0 and 1, so that d reduces the probability of subsequent infection over time interval Δt . After the previous infection is cleared at time t_{clr} , immunity wanes at rate w .

- (iii) In the additional risk model, the risk of an initial infection is determined as in the memoryless model (Equation 1), but the risk of subsequent infection is allowed to be higher. The force of infection thus includes a nonspecific additive risk factor, which is distinct from the effects of behavioral and demographic risk factors (Equation 4, Figure 2):

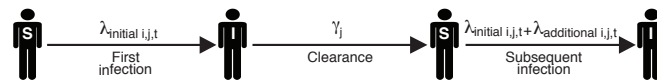


Figure 2: Schematic showing the dynamics for one individual host i and one HPV type j . The labels "S" and "I" indicate the susceptible and infected states, respectively. Before the initial infection, individual i 's per-capita, per-unit-time infection risk with type j , λ_{ijt} , is determined by the type-specific baseline force of infection and the effect of host-specific risk factors, as in Equation 1. The duration of each infection is drawn from a Gamma distribution with mean $1/\gamma_j$. Previous infection modulates the risk of a subsequent infection as in Equation 4.

Table 1: Comparison of Candidate Models

Type	Model (n parameters)	Log likelihood	SE	ΔAICc
HPV62	Memoryless (13)	-2026.1	1.4	389.6
	Homologous immunity (15)	-2027.4	1.5	396.3
	Additional risk (17)	-1827.2	0.4	0
HPV16	Memoryless (13)	-1663.2	1.1	117.9
	Homologous immunity (15)	-1663.7	0.7	123.1
	Additional risk (17)	-1600.1	0.1	0
HPV89	Memoryless (13)	-1902.3	1.3	137.4
	Homologous immunity (15)	-1902.2	1.5	141.2
	Additional risk (17)	-1829.5	0.1	0
HPV51	Memoryless (13)	-1670.8	0.6	63.9
	Homologous immunity (15)	-1670.4	1.7	67.3
	Additional risk (17)	-1634.7	0.7	0
HPV84	Memoryless (13)	-1979.6	1.7	270.3
	Homologous immunity (15)	-1980.5	1.1	276.3
	Additional risk (17)	-1840.3	0.4	0
HPV6	Memoryless (13)	-1461.0	1.0	121.7
	Homologous immunity (15)	-1461.5	1.3	126.8
	Additional risk (17)	-1396.1	0.6	0

$$\lambda_{ijt} = \lambda_{0j} f(\vec{\alpha}_j^T \mathbf{C}_{it}) + I_{\text{prev}} d_{jc_i} e^{-w_j(t-t_{\text{clr}})}. \quad (4)$$

Here the additional risk parameter d_{jc_i} depends on the infecting type j , as well as the sexual subclass c_i of host i . The variable I_{prev} indicates whether there was a previous infection with type j .

We also included an observation model that determines the probability of the observed data using the known sensitivity and specificity of the HPV genotyping test. We fitted models to the data using a likelihood-based nonlinear fitting routine [49, 48]. To choose the model that best explains the data, we used the corrected Akaike Information Criterion (AICc), which balances the better fit of more complex models against the parsimony of simpler models, while including a correction for finite sample size [47]. The best model has the smallest AICc value.

We first tested whether infection with a particular HPV type reduces the risk of subsequent infection with the same type, allowing for variation in the degree and duration of protective homologous immunity (Equation 3). The homologous immunity model performed consistently worse than the memoryless model (Table 1), such that the additional parameters for homologous immunity increased model complexity without improving the fit, resulting in higher AICc scores (Table 1). There is therefore no evidence for homologous immunity against any of the six examined HPV types, suggesting that intra-type competition is weak to nonexistent.

Because separate types are less closely related, competition between types should be weaker still. Previous work has nevertheless speculated that there may be cross-immunity between HPV types. In particular, virus-like particles of HPV16 have been shown to induce a low level of neutralizing antibodies against HPV31, and clinical trial data suggest that current vaccines, which immunize against HPV16, provide partial protection against HPV31 [21, 55]. We therefore tested for competition between HPV16 and HPV31 by fitting a model in which previous infection with HPV31 affects the risk of infection with HPV16 (Supporting Information 2.2). Our estimate of the relative risk of HPV16 infection given previous HPV31 infection is centered around 1 (1.3, 95% C.I.: [0.5, 2.0]), suggesting that the two types do not compete. This lack of an effect may be partly due to the low prevalence of HPV31 (Figure 1) and correspondingly low statistical power, but the lack of even a trend toward competition nevertheless suggests that there is no interaction. Accordingly, neither intra-type nor inter-type competition has strong effects on HPV dynamics.

Past infection with HPV strongly increases the risk of future infection with the same type

The additional risk model fits the data vastly better than the homologous immunity and memoryless models (Table 1) for all 6 HPV types. The additional risk could reflect repeated exposure to the same HPV type by sexual activity with infected partners. To control for this possibility, we fitted the magnitude of the additional risk d separately for three different sexual subclasses: individuals reporting no recent sexual activity (celibate individuals), individuals reporting one recent sexual partner, and individuals reporting multiple recent sexual partners. In this analysis, we included only people who remained in the same sexual subclass for at least three consecutive years. For all six HPV types, previous infection significantly raises the risk of subsequent homologous infection irrespective of sexual subclass (Figure 3A).

The high additional risk experienced by celibate individuals (d_{celibate}) in particular strongly suggests that serial infections are driven by factors other than sexual transmission. Crucially, our estimate of the additional risk is uncorrelated with our estimate of baseline infection risk, λ_{0j} (Figure 3B), showing that the high estimated additional risk does not reflect statistical non-identifiability between additional risk and baseline risk. Estimates of the additional risk between different sexual subclasses have the same magnitude (Figure 3A), demonstrating that the additional risk varies only slightly across sexual subclasses. To quantify the impact of previous infection, we calculated the effect of previous infection on the total risk of a subsequent homologous infection at $t = 0, 1$, and 3 years after clearing the previous infection. For at least several years after the initial infection is cleared, the additional risk due to previous infection accounts for more than 90% of the force of infection (Figure S2). Moreover, immediately following a previous infection, the one-year probability of re-infection with HPV16 (Equation 5) is on average 20.4-fold higher than the probability of infection in a previously uninfected individual. The average increase is 19.1 - 20.5-fold among HPV types. At $t = 3$ years after clearing the previous infection, the one-year probability of reinfection with HPV16 remains 13.5-fold higher (7.4 - 20.5-fold among types).

The finding that initial infection strongly increases the risk of subsequent infection independently of sexual subclass is consistent with two major biological explanations: repeat infections, presumably due to auto-inoculation between anatomical sites, or episodic reactivation of latent infection. Because our models take into account 11 different risk factors besides previous infection, it is unlikely that the increased risk instead represents confounding by unmeasured host risk factors. Completely ruling out such unmeasured covariates is ultimately impossible without performing controlled experiments. As a preliminary approach to test for confounding, we repeated our estimation of the additional risk parameter d using a model that includes additional measured covariates for a total of 17 host-specific risk factors. In this more complex model, the additional risk d still accounts for more than 90% of the force of infection following clearance of the initial infection, and the model with more covariates did not provide a better fit to the data than the best-fit model with 11 covariates (Supporting Information 2.3). This result suggests that confounding by sexual risk factors is unlikely, while emphasizing the inability of traditional risk factors to explain the vast majority of HPV infections.

Modest differences in host-specific risk factors suggest ecological differences between HPV types and highlight high-risk subpopulations

Although the additional risk conferred by past infection is substantial, a model without host-specific risk factors fits the data far worse than the full additional-risk model (Equation 4) for all types (Supporting Information 2.1). Moreover, the effects of these host-specific risk factors vary among HPV types. To understand the effects of this variation, we inserted our estimates of the baseline force of infection and the covariate effects for each HPV type into the best-fit model to calculate the distribution of infection risk in individuals who have never been infected. The expected time to infection ($1/\lambda_{ijt}$), a measure of infection risk, is generally low but varies by orders of magnitude among individuals in the naive population (Figure S1), suggesting that the risk of initial infection with any type is concentrated among a few high-risk individuals.

Furthermore, the effects of each covariate show interesting similarities and differences between types (Figure 4). Increases in the number of recent female partners strongly increases risk for all HPV types, emphasizing the importance of heterosexual transmission. The addition of a single female sexual partner raises instantaneous infection risk by 80% to 120% among types compared to the risk in individuals with no recent female partners. The number of male partners, however, has divergent effects. The addition of a single male sexual partner *reduces* instantaneous risk with HPV16 by 50%, but *increases* the risk of infection with types HPV84 and HPV89 by 20% and 70%, respectively. Most other covariates were significant for only some of the types, indicating subtle differences between types. Correlations in risk factors between types were not associated with oncogenicity, such that HPV16 and HPV51, the two oncogenic types, were not more closely associated with one another based on their covariate profiles than they were to other types (Figure 4). Phylogenetic classification similarly did not reflect meaningful groupings of HPV types by their covariate profiles: HPV89, HPV62, and HPV84, all from the α -3 species, were not more closely associated to one another than to other types (Figure 4).

Our estimates of the effects of type-specific risk factors are generally consistent with previous studies, but there are some

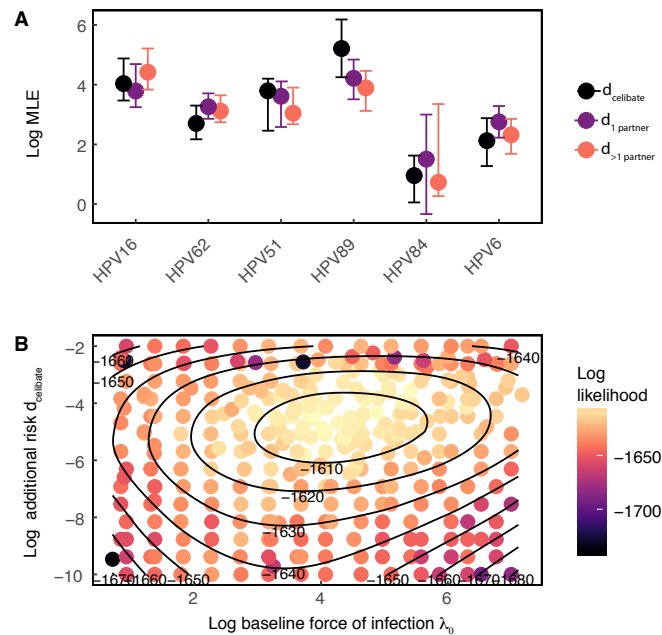


Figure 3: Inference of the additional risk d by sexual subclass. **A** Log estimates of d by sexual subclass for each HPV type. Points and error bars give the maximum likelihood estimate and 95% confidence interval, respectively. **B** Bivariate likelihood profile for the additional risk d_{celibate} and the baseline infection risk λ_0 (shown on a log scale). Results shown for HPV16.

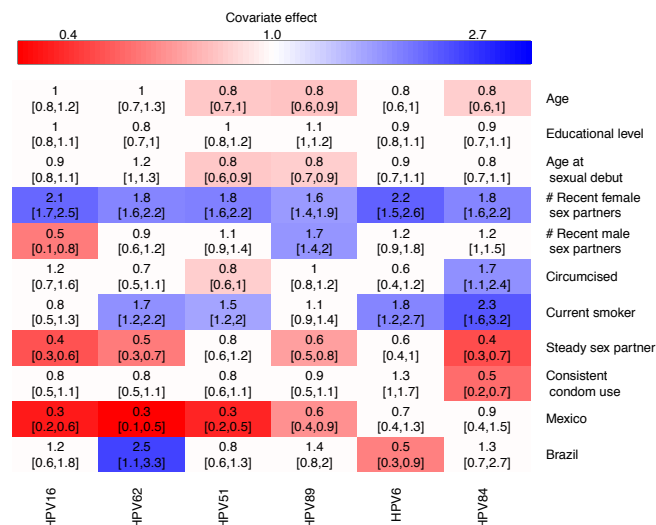


Figure 4: Inferred effect of host-specific covariates on the force of infection for each HPV type. Values provide the maximum likelihood parameter estimate and the 95% confidence interval. Cells colored in red denote statistically significant negative effects, and cells colored in blue denote statistically significant positive effects. HPV types are arranged by complete-linkage hierarchical clustering [63, 27].

exceptions. First, a previous analysis of the HIM data showed no difference in HPV16 incidence between men who have sex with men and men who have sex with women [85], but our results instead show that risk of HPV16 infection declines with an increase in the number of male sex partners, suggesting that HPV16 infection depends largely on heterosexual transmission. Meanwhile, our results show that the risk of infection with HPV51, HPV89, HPV6, and HPV84 increases with the number of male or female partners, consistent with previous work, although the effect was statistically significant only for HPV89. Finally, the previous analysis concluded that circumcision increases the risk of HPV51 [1], but our results suggest that circumcision instead reduces the risk of HPV51.

Several features of our modeling approach, however, may provide greater statistical power than previous analyses. First, our models account for variation in the risk per unit time, and they describe the dynamics in a way that is unconstrained and unbiased by the frequency of observations. Second, we infer the contribution of host-specific risk factors to the infection risk separately from the dynamics of previous infection. We therefore believe that our results are robust and that differences with respect to previous analyses arise from differences in the model structure.

Discussion

Understanding the dynamics of HPV transmission is crucial to explain the coexistence of over 200 low-prevalence HPV types and to accurately predict the impact of multivalent HPV vaccines. We find little evidence of competition within or between HPV types, and instead we find high rates of reinfection or persistence by the same type. Measured epidemiological risk factors do explain at least some of the risk of initial infection, but they explain less than 10% of the risk after a man has recovered from his first infection. Previous infection with a particular type in contrast increases the risk of infection with the same type by 7- to 21-fold even three years after clearance. The inferred lack of both homologous and heterologous immunity implies that HPV type diversity cannot be explained by negative frequency-dependent selection, which promotes the coexistence of other immunologically distinct pathogen strains. The high prevalence of HPV instead results from continuous or repeated infection with multiple, apparently independent virus types. These types differ slightly in their risk factors for transmission to males, suggesting that ecological niche partitioning could play a modest role in further promoting coexistence.

Our results add to growing evidence that HPV infection, as opposed to vaccination, does not confer protective homologous immunity in men. Although we allowed the strength and duration of immunity to vary following initial infection, we could not detect any protection against reinfection with the same type. Because sterilizing immunity leads to lower overall infection rates, previous models that assume that infection results in protective immunity in men likely underestimated vaccine effectiveness [105, 24, 80, 7, 116, 11, 4, 26, 30, 61, 117, 5]. The well known effectiveness of the HPV vaccine may therefore be even higher than previously believed.

Our conclusion that there is no homologous immunity in men supports the hypothesis that genital infection differs between men and women. Although the durations and distributions of types in genital HPV infection are similar in men and women [44, 37], the prevalence of genital HPV is higher in men [44]. Acquired immunity has been proposed to explain declining cervical HPV prevalence by age in women in some countries [10, 102], whereas HPV prevalence does not change with age in men [34, 44]. One model showed that acquired immunity is necessary to explain age-specific patterns of prevalence in women [76], in contrast to our results in men. Furthermore, the seroprevalence of some types is higher in women than in men from the same source population [73, 17]. However, homologous protection in women is still likely to be weak [120, 6], supporting the idea that limited competition contributes to high HPV prevalence.

The high risk of recurring infection is consistent with either auto-inoculation across anatomic sites or with episodic reactivation of latent infection. The high risk of reinfection in celibate individuals makes confounding due to infected but unreported partners unlikely (Figure 3). The importance of auto-inoculation is supported by studies showing type-concordant HPV infection across anatomic sites. First, HPV DNA has been detected on the fingers of patients with genital warts [106], suggesting that HPV is transmitted between the hands and the genitals. Second, an analysis of a subcohort of the HIM study demonstrated a 1.5 to 15-fold increase in the risk of anal HPV infection after genital infection [85], and a similar analysis of sexually active women in Hawaii found a 20-fold increase in the risk of recurrent anal HPV infection given infection at the cervix [38]. Moreover, 63% of the cases of secondary anal HPV in the Hawaiian study occurred without a self-reported history of anal sex. Significantly, the magnitude of the effect in each of these studies is consistent with our estimate that previous infection leads to a 20-fold increase in the risk of subsequent infection.

Apparent reinfection could instead be due to the reactivation of an infection from a reservoir of latent virus. Whether HPV persists in a latent viral state remains unknown [100, 41], but an animal model of cottontail oral papillomavirus provides evidence of a latent viral reservoir in the basal layer of the mucosal epithelium [72]. HIV-positive women who are sexually inactive have a higher risk of recurrent HPV infection compared to HIV-negative controls [113], and among HIV-positive females, higher CD4+ T cell counts

are negatively correlated with recurrent infection risk [110]. The latter two studies in particular have been interpreted as evidence of reactivation of latent infection [39], but either effect may instead result from increases in susceptibility to auto-infection rather than to diminished suppression of latent infection. Furthermore, the studies in HIV-positive individuals suggest that reactivation may require suppression of cellular immunity, and may therefore be rare in healthy individuals.

Vaccine efficacy trials furthermore suggest that reinfection is more common than reactivation. Women with previous HPV-related disease who received the quadrivalent vaccine were protected against HPV-related lesions [51], suggesting that reinfection is more common than reactivation. Similarly, one study showed 100% efficacy of the HPV vaccine against HPV-related disease in individuals with serological evidence of past HPV 6, 11, 16, or 18 infection [87]. Because HPV vaccines prevent infection by inducing antibodies that block viral entry at the epithelial basement membrane, such antibodies would likely not prevent reactivation of latent virus. If indeed vaccines do not affect the disease course of reactivated infections, then these vaccine efficacy trials suggest that the vaccine diminishes the incidence of HPV lesions by blocking true reinfection.

In short, although our estimate of the high additional risk from previous infection allows for the possibility that subsequent infections are due to the combined effects of reinfection and reactivation, other evidence suggests that autoinfection is more likely. This is important for vaccination policies, which often target people who are young enough that they are not yet sexually active. Vaccinating young people clearly reduces infection rates [74, 75], but if vaccination also reduces the risk of autoinfection, then vaccinating previously infected individuals may also strongly reduce HPV prevalence.

The type-specific effects of demographic and behavioral risk factors that we inferred suggest that modest differences exist in the host subpopulations supporting each HPV type. These differences highlight functional ecological distinctions between types that may further contribute to type coexistence. For all types, the major determinant of infection risk is the number of recent female sex partners, suggesting a central role for heterosexual transmission. Additionally, current smoking increases risk in most types, consistent with other studies [33, 35, 44, 91]. Smoking can suppress mucosal and cellular immunity [62], but its effects may be confounded with other potentially unobserved, high-risk behaviors. Although shared risk factors account for most of the initial infection risk for each type, there are important distinctions too. For instance, the effect of circumcision on the risk of HPV infection is type-specific. Several randomized, controlled trials of male circumcision have demonstrated that circumcision protects against any HPV infection [3, 42], whereas observational studies have instead documented either no effect or increased risk [44, 68]. These conflicting results may have arisen from confounding by differences between types. Foreskin consists of unkeratinized mucosal epithelium, whereas the exposed tissue in circumcised men is made up of stratified squamous epithelium [42]. Our finding that circumcision increases the risk of infection with some HPV types (e.g. HPV84) and decreases the risk with others (e.g. HPV51) could reflect the adaptation of HPV types to different epithelia in the male genital tract.

Our modeling approach has several limitations, notably that we cannot distinguish between reactivation and auto-infection, and that we cannot rule out the possibility that unmeasured covariates affected our results. In addition, however, our model of homologous immunity assumes that protection arises immediately after infection, but protection may be lagged [80, 101]. The poor fit of the model that includes protective immunity nevertheless suggests that any such effects are quite weak, but allowing for more complex forms of homologous immunity is an important next step. Likewise, we assume that our estimate of the baseline infection risk for any type j , λ_{0j} , captures the baseline risk of infection in naive hosts, but we have no information on HPV exposure or changes in risk factors that occurred before the study, nor do we have information on genetic differences in host susceptibility. Finally, the data track HPV infection in the genitals, ignoring other sites.

These caveats notwithstanding, the strong effect of initial infection on subsequent infection risk that we inferred is important for the design of epidemiological studies and models to inform public health policy. The effects of recurrent infection on subsequent risk suggest that vaccination may alter overall infection rates in ways that are not predicted from standard models of HPV dynamics [89, 24, 96]. The HPV vaccine therefore has the potential to lower HPV prevalence far more than previously expected.

Materials and Methods

Data

The HPV in Men (HIM) study is a multinational cohort study of HPV infection unvaccinated men aged 18-70 years. The HIM study provides the infection status with 37 types of HPV, confirmed by Linear Array PCR test [40], as well as host demographic and behavioral information. The study enrolled 4,123 participants between 2005 and 2009 and tracked men longitudinally over five years. Men were recruited from three cities: Tampa, Florida, USA; Cuernavaca, Mexico; and Sao Paulo, Brazil. Detailed study methods are described elsewhere [34, 35]. We excluded individuals that failed to meet the full eligibility criteria described by the HIM study [34], which included no prior diagnosis of genital cancer, warts, HIV, or other STIs. We classified men at each visit by sexual subclass according to their recent number of sexual partners, defined as the total number of reported sexual partners within the last six

months. We defined three sexual subclasses: individuals that reported no sexual activity in the last six months (celibate individuals), individuals that reported one recent sex partner, and individuals that reported multiple (two or more) recent sex partners. The data were limited by individuals who switched between sexual subclasses. We included only the $n = 1,099$ individuals that remained in their subclass for at least 3 years.

The data for each individual consist of binary time series describing infection status with each HPV type over a maximum of 10 clinic visits. The median number of visits among included participants was 10 visits. Data about host-specific covariates were drawn from a risk factor questionnaire administered at every visit. Additional characteristics of the data and a description of the covariate variables are provided in the Supporting Information (section 1).

Model of HPV Dynamics

Our models are two-state discrete time partially observed Markov processes (POMPs), in which an individual is either infected (1) or uninfected (0) at any time with an HPV type. The $0 \rightarrow 1$ and $1 \rightarrow 0$ transitions are unobserved. The probability of infection (a $0 \rightarrow 1$ transition) is determined by the force of infection, λ . Infection of individual i with HPV type j occurs at rate λ_{ijt} , and the time to infection is thus exponentially distributed. Thus, the probability of infection of individual i with type j (a $0 \rightarrow 1$ transition) over time interval Δt is given by

$$p(1_{t+\Delta t}|0_t) = 1 - e^{-\lambda_{ijt}\Delta t}. \quad (5)$$

The measurement model is a Bernoulli distribution that relates each observation Y_{ijt} in an individual's binary time series to the latent state X_{ijt} simulated under the model,

$$Y_{ijt} \sim \text{Bernoulli}(X_{ijt}(1 - P_{\text{FP}}) + (1 - X_{ijt})P_{\text{FN}}), \quad (6)$$

where Y_{ijt} is the observation of individual i with type j at time t , X_{ijt} is the corresponding latent state, P_{FP} is the rate of false positives, and P_{FN} is the rate of false negatives.

Table S3 describes the parameters for each candidate model. Each model was fit separately for each HPV type. Time-varying covariates (age, recent numbers of male and female sex partners, the presence of a steady sex partner, smoking status, educational level, and condom use) were updated at each visit. We parsimoniously used a log-linear model to model the covariate effects (Equation 2), such that each covariate effect α is directly interpretable as a multiplicative effect on the instantaneous risk of infection λ_{ijt} . To identify the force of infection separately from the infection duration, the duration of infection with each HPV type j was drawn from a gamma distribution, where the shape (k_j) and scale (θ_j) were fixed to match the empirical distribution of infection durations in the dataset for type j (Table S4, Figure S7). The rate of false positives and false negatives in the measurement model were fixed based on the sensitivity and specificity of the genotyping test.

Likelihood-based inference

Inference by maximum likelihood was carried out using multiple iterated filtering (MIF) [48] implemented in the R package “panelPomp” (version 0.3.1), which extends methods of the “pomp” package to longitudinal data [59, 58]. Briefly, iterated filtering is an algorithm that uses sequential Monte Carlo (SMC) to approximate maximum likelihood estimates of parameters from POMP models. SMC uses a population of particles drawn from the parameters of a given model to generate Monte Carlo samples of the latent dynamic variables and evaluate the likelihood of observed time series [2, 20]. Iterated filtering successively filters the particle population, perturbing the parameters between iterations. The perturbations decrease in amplitude over time, allowing convergence at the maximum likelihood estimate.

The extension of SMC and iterated filtering to longitudinal panel data has been previously described [97], and we extended longitudinal POMP methods to binary data. The data for each HPV type is a set of binary time series, or panel units, describing the observed infection trajectory for each individual. The panel POMP contains a POMP model for each individual, and individuals share parameters. To evaluate the likelihood of a shared parameter set, SMC is carried out over the time series for each individual to generate a unit log-likelihood. The log likelihood of the panel POMP object is the sum of the individuals' log likelihoods. All optimization routines were carried out using 20,000 particles to overcome high Monte Carlo error (Supporting Information 3).

For each model, we initialized the iterated filtering with 100 random parameter combinations. Optimization involved series of successive MIF searches, with the output of each search serving as the initial conditions for the subsequent search. The likelihood of the output for each search was calculated by averaging the likelihood from ten passes through the particle filter using 50,000 particles. The optimization was repeated until additional operations did not arrive at a higher maximum likelihood (Supporting Information 3).

For model selection, we used the corrected Akaike Information Criterion (AICc) [47]. We obtained parameter estimates (maximum likelihood estimates and 95% confidence intervals) by constructing likelihood profiles. We used Monte Carlo Adjusted Profile methods [49] to obtain a smoothed estimate of the profile that accounts for the increased Monte Carlo error associated with longitudinal data. The lower and upper limits of the 95% confidence interval were the points that lay 1.92 log-likelihood units below the maximum likelihood estimate on the smoothed profile curve (the points corresponding to one-half the 95% critical value for a χ^2 distribution with one degree of freedom).

Data accessibility

The data and code for this analysis are available at <https://github.com/cobeylab/HPV-model>.

Author Contributions

S.R., S.C., and G.D. designed the research; S.R. performed research and analysis; E.B. provided computational expertise; V.D. provided statistical expertise; A.G., L.V., and E.L. provided data; A.G. provided clinical expertise; S.R., S.C., and G.D. wrote the manuscript.

Acknowledgements

Ken Alexander was instrumental in the initiation of this project. We thank Daniel Zinder for helpful discussions about the model construction. This work was completed with resources provided by the University of Chicago Research Computing Center and with funding provided by the National Institute of Allergies and Infectious Diseases (NIH F30AI124636 to S.R.) and by the University of Chicago Medical Scientist Training Program (NIH T32GM007281).

References

- [1] G. Albero, X. Castellsagué, H.-Y. Lin, W. Fulp, L. L. Villa, E. Lazcano-Ponce, M. Papenfuss, M. Abrahamsen, J. Salmerón, M. Quiterio, A. G. Nyitray, B. Lu, F. X. Bosch, and A. R. Giuliano. Male circumcision and the incidence and clearance of genital human papillomavirus (HPV) infection in men: the HPV Infection in men (HIM) cohort study. *BMC Infectious Diseases*, 14(1):75, 2014.
- [2] M. Arulampalam and S. Maskell. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.
- [3] B. Auvert, J. Sobngwi-Tambekou, E. Cutler, M. Nieuwoudt, P. Lissouba, A. Puren, and D. Taljaard. Effect of male circumcision on the prevalence of high-risk human papillomavirus in young men: results of a randomized controlled trial conducted in Orange Farm, South Africa. *Journal of Infectious Diseases*, 199(1):14–19, 1 2009.
- [4] R. V. Barnabas, P. Laukkanen, P. Koskela, O. Kontula, M. Lehtinen, and G. P. Garnett. Epidemiology of HPV 16 and Cervical Cancer in Finland and the Potential Impact of Vaccination: Mathematical Modelling Analyses. *PLoS Medicine*, 3(5):e138, 4 2006.
- [5] I. Baussano, G. Garnett, N. Segnan, G. Ronco, and P. Vineis. Modelling patterns of clearance of HPV-16 infection and vaccination efficacy. *Vaccine*, 29(6):1270–7, 3 2011.
- [6] D. C. Beachler, G. Jenkins, M. Safaeian, A. R. Kreimer, and N. Wentzensen. Natural Acquired Immunity Against Subsequent Genital Human Papillomavirus Infection: A Systematic Review and Meta-analysis. *Journal of Infectious Diseases*, 213(9):1444–1454, 5 2016.
- [7] J. A. Bogaards, M. Kretzschmar, M. Xiridou, C. J. L. M. Meijer, J. Berkhof, and J. Wallinga. Sex-Specific Immunization for Sexually Transmitted Infections Such as Human Papillomavirus: Insights from Mathematical Models. *PLoS Medicine*, 8(12):e1001147, 12 2011.
- [8] J. a. Bogaards, M. Xiridou, V. M. H. Coupé, C. J. L. M. Meijer, J. Wallinga, and J. Berkhof. Model-based estimation of viral transmissibility and infection-induced resistance from the age-dependent prevalence of infection for 14 high-risk types of human papillomavirus. *American Journal of Epidemiology*, 171(7):817–25, 4 2010.

- [9] B. Bolker. Likelihood and all that. In *Ecological Models and Data in R*, chapter 6, pages 169–221. Princeton University Press, 508 edition, 2007.
- [10] F. X. Bosch, X. Castellsagué, and S. de Sanjosé. HPV and cervical cancer: screening or vaccination? *British Journal of Cancer*, 98(1):15–21, 2008.
- [11] M. Brisson, A. Bénard, M. Drolet, M.-c. Boily, H. C. Turner, M. Brisson, A. Bénard, M. Drolet, J. A. Bogaards, I. Baussano, S. Vänskä, M. Jit, M.-C. Boily, M. A. Smith, J. Berkhof, K. Canfell, H. W. Chesson, E. A. Burger, Y. H. Choi, B. Freiesleben De Blasio, S. J. De Vlas, G. Guzzetta, J. A. C Hontelez, J. Horn, M. R. Jepsen, J. J. Kim, F. Lazzarato, S. M. Matthijsse, R. Mikolajczyk, A. Pavelyev, M. Pillsbury, L. Anne Shafer, S. P. Tully, H. C. Turner, C. Usher, and C. Walsh. Population-level impact, herd immunity, and elimination after human papillomavirus vaccination: a systematic review and meta-analysis of predictions from transmission-dynamic models. *Funding Canadian Institutes of Health Research*, 1(16):8–17, 2016.
- [12] M. Brisson, N. Van de Velde, and M.-C. Boily. Different population-level vaccination effectiveness for HPV types 16, 18, 6 and 11. *Sexually Transmitted Infections*, 87(1):41–3, 2011.
- [13] C. B. Buck, P. M. Day, and B. L. Trus. The papillomavirus major capsid protein L1. *Virology*, 445(1-2):169–74, 10 2013.
- [14] A. N. Burchell, R. L. Winer, S. de Sanjosé, and E. L. Franco. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. *Vaccine*, 24(Supplement 3):52–61, 8 2006.
- [15] A. K. Chaturvedi. Prevalence and Clustering Patterns of Human Papillomavirus Genotypes in Multiple Infections. *Cancer Epidemiology, Biomarkers & Prevention*, 14(10):2439–2445, 2005.
- [16] A. K. Chaturvedi, H. A. Katki, A. Hildesheim, A. C. Rodríguez, W. Quint, M. Schiffman, L. J. Van Doorn, C. Porras, S. Wacholder, P. Gonzalez, M. E. Sherman, and R. Herrero. Human papillomavirus infection with multiple types: Pattern of coinfection and risk of cervical disease. *Journal of Infectious Diseases*, 203(7):910–920, 2011.
- [17] G. M. Clifford, H.-R. Shin, J.-K. Oh, T. Waterboer, Y.-H. Ju, S. Vaccarella, W. Quint, M. Pawlita, and S. Franceschi. Serologic Response to Oncogenic Human Papillomavirus Types in Male and Female University Students in Busan, South Korea. *Cancer Epidemiology Biomarkers & Prevention*, 16(9):1874–1879, 9 2007.
- [18] S. Cobey. Pathogen evolution and the immunological niche. *Annals of the New York Academy of Sciences*, 1320(1):1–15, 7 2014.
- [19] E.-M. de Villiers, C. Fauquet, T. R. Broker, H.-U. Bernard, and H. zur Hausen. Classification of papillomaviruses. *Virology*, 324(1):17–27, 6 2004.
- [20] A. Doucet, N. d. Freitas, and N. Gordon. Sequential Monte Carlo Methods in Practice. Series Statistics For Engineering and Information Science, 2001.
- [21] E. Draper, S. Bissett, R. Howell-Jones, and D. Edwards. Neutralization of non-vaccine human papillomavirus pseudoviruses from the A7 and A9 species groups by bivalent HPV vaccine sera. *Vaccine*, 29(47):8585–8590, 2011.
- [22] F. Drasgow, Drasgow, and Fritz. Polychoric and Polyserial Correlations. In *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 9 edition, 8 2006.
- [23] E. Dunne, C. Nielson, and K. Stone. Prevalence of HPV infection among men: a systematic review of the literature. *Journal of Infectious Diseases*, 194(9):1044–1057, 2006.
- [24] D. P. Durham, M. L. Ndeffo-Mbah, L. A. Skrip, F. K. Jones, C. T. Bauch, and A. P. Galvani. National- and state-level impact and cost-effectiveness of nonavalent HPV vaccination in the United States. *Proceedings of the National Academy of Sciences*, 113(18):5107–5112, 2016.
- [25] D. P. Durham, E. M. Poolman, Y. Ibuka, J. P. Townsend, and A. P. Galvani. Reevaluation of epidemiological data demonstrates that it is consistent with cross-immunity among human papillomavirus types. *Journal of Infectious Diseases*, 206(8):1291–1298, 2012.
- [26] E. H. Elbasha, E. J. Dasbach, and R. P. Insinga. Model for Assessing Human Papillomavirus Vaccination Strategies. *Emerging Infectious Diseases*, 13(1):28–41, 1 2007.

- [27] B. Everitt, S. Landau, and M. Leese. Hierarchical Clustering. In *Clustering Analysis*, chapter 4, pages 71–110. John Wiley & Sons Ltd., 5 edition, 2001.
- [28] S. Franceschi and I. Baussano. Naturally Acquired Immunity Against Human Papillomavirus (HPV): Why It Matters in the HPV Vaccine Era. *Journal of Infectious Diseases*, 210(4):507–509, 2014.
- [29] I. H. Frazer. Interaction of human papillomaviruses with the host immune system: A well evolved relationship. *Virology*, 384(2):410–414, 2009.
- [30] K. M. French, R. V. Barnabas, M. Lehtinen, O. Kontula, E. Pukkala, J. Dillner, and G. P. Garnett. Strategies for the introduction of human papillomavirus vaccination: modelling the optimum age- and sex-specific pattern of vaccination in Finland. *British Journal of Cancer*, 96(3):514–518, 2 2007.
- [31] R. G. Fruchter, M. Maiman, A. Sedlis, L. Bartley, L. Camilien, and C. D. Arrastia. Multiple recurrences of cervical intraepithelial neoplasia in women with the human immunodeficiency virus. *Obstetrics and Gynecology*, 87(3):338–344, 3 1996.
- [32] M. L. Gillison, W. M. Koch, R. B. Capone, M. Spafford, W. H. Westra, L. Wu, M. L. Zahurak, R. W. Daniel, M. Viglione, D. E. Symer, K. V. Shah, and D. Sidransky. Evidence for a Causal Association Between Human Papillomavirus and a Subset of Head and Neck Cancers. *Journal of the National Cancer Institute*, 92(9):709–720, 5 2000.
- [33] A. R. Giuliano, E. Lazcano, L. L. Villa, R. Flores, J. Salmeron, J.-H. Lee, M. Papenfuss, M. Abrahamsen, M. L. Baggio, R. Silva, and M. Quiterio. Circumcision and sexual behavior: factors independently associated with human papillomavirus detection among men in the HIM study. *International Journal of Cancer*, 124(6):1251–7, 3 2009.
- [34] A. R. Giuliano, E. Lazcano-Ponce, L. L. Villa, R. Flores, J. Salmeron, J.-H. Lee, M. R. Papenfuss, M. Abrahamsen, E. Jolles, C. M. Nielson, M. L. Baggio, R. Silva, and M. Quiterio. The human papillomavirus infection in men study: human papillomavirus prevalence and type distribution among men residing in Brazil, Mexico, and the United States. *Cancer Epidemiology, Biomarkers & Prevention*, 17(8):2036–43, 8 2008.
- [35] A. R. Giuliano, J.-H. Lee, W. Fulp, L. L. Villa, E. Lazcano, M. R. Papenfuss, M. Abrahamsen, J. Salmeron, G. M. Anic, D. E. Rollison, and D. Smith. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. *Lancet*, 377(9769):932–40, 3 2011.
- [36] A. R. Giuliano, A. G. Nyitray, A. R. Kreimer, C. M. Pierce Campbell, M. T. Goodman, S. L. Sudenga, J. Monsonego, and S. Franceschi. EUROGIN 2014 roadmap: Differences in human papillomavirus infection natural history, transmission and human papillomavirus-related cancer incidence by gender and anatomic site of infection. *International Journal of Cancer*, 136(12):2752–2760, 2015.
- [37] A. R. Giuliano, A. G. Nyitray, A. R. Kreimer, C. M. Pierce Campbell, M. T. Goodman, S. L. Sudenga, J. Monsonego, and S. Franceschi. EUROGIN 2014 roadmap: Differences in human papillomavirus infection natural history, transmission and human papillomavirus-related cancer incidence by gender and anatomic site of infection. *International Journal of Cancer*, 136(12):2752–2760, 2015.
- [38] M. T. Goodman, Y. B. Shvetsov, K. McDuffie, L. R. Wilkens, X. Zhu, P. J. Thompson, L. Ning, J. Killeen, L. Kamemoto, and B. Y. Hernandez. Sequential acquisition of human papillomavirus (HPV) infection of the anus and cervix: the Hawaii HPV Cohort Study. *Journal of Infectious diseases*, 201(9):1331–9, 5 2010.
- [39] P. E. Gravitt. Evidence and impact of human papillomavirus latency. *Open Virology Journal*, 6:198–203, 2012.
- [40] P. E. Gravitt, C. L. Peyton, R. J. Apple, and C. M. Wheeler. Genotyping of 27 human papillomavirus types by using L1 consensus PCR products by a single-hybridization, reverse line blot detection method. *Journal of Clinical Microbiology*, 36(10):3020–7, 10 1998.
- [41] P. E. Gravitt, A. Schneider, A. Kaufmann, F. Bosch, R. Clark, and T. Kupper. The known unknowns of HPV natural history. *Journal of Clinical Investigation*, 121(12):4593–4599, 12 2011.

- [42] R. H. Gray, D. Serwadda, X. Kong, F. Makumbi, G. Kigozi, P. E. Gravitt, S. Watya, F. Nalugoda, V. Ssempijja, A. A. R. Tobian, N. Kiwanuka, L. H. Moulton, N. K. Sewankambo, S. J. Reynolds, T. C. Quinn, B. Iga, O. Laeyendecker, A. E. Oliver, and M. J. Wawer. Male circumcision decreases acquisition and increases clearance of high-risk human papillomavirus in HIV-negative men: a randomized trial in Rakai, Uganda. *Journal of Infectious Diseases*, 201(10):1455–62, 5 2010.
- [43] S. Gupta and R. Anderson. Population Structure of Pathogens: The Role of Immune Selection. *Parasitology Today*, 15(12):497–501, 12 1999.
- [44] J. J. Han, T. H. Beltran, J. W. Song, J. Klaric, and Y. S. Choi. Prevalence of Genital Human Papillomavirus Infection and Human Papillomavirus Vaccination Rates Among US Adult Men. *JAMA Oncology*, 3(6):810–816, 2017.
- [45] S. Hartwig, J. J. Baldauf, G. Dominiak-Felden, F. Simondon, L. Alemany, S. de Sanjosé, and X. Castellsagué. Estimation of the epidemiological burden of HPV-related anogenital cancers, precancerous lesions, and genital warts in women and men in Europe: Potential additional benefit of a nine-valent second generation HPV vaccine compared to first generation HPV vaccines. *Papillomavirus Research*, 1:90–100, 2015.
- [46] H. z. Hausen. Papillomavirus infections – a major cause of human cancers. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1288(2):F55–F78, 10 1996.
- [47] C. M. Hurvich and C.-L. Tsai. Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76(2):297, 6 1989.
- [48] E. L. Ionides, C. Bretó, and A. A. King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443, 2006.
- [49] E. L. Ionides, C. Breto, J. Park, R. A. Smith, and A. A. King. Monte Carlo profile confidence intervals for dynamic systems. *Journal of The Royal Society Interface*, 14(132):20170126, 2017.
- [50] H. C. Johnson, K. M. Elfström, and W. J. Edmunds. Inference of Type-Specific HPV Transmissibility, Progression and Clearance Rates: A Mathematical Modelling Approach. *PLoS ONE*, 7(11):1–13, 2012.
- [51] E. A. Joura, S. M. Garland, J. Paavonen, D. G. Ferris, G. Perez, K. A. Ault, W. K. Huh, H. L. Sings, M. K. James, and R. M. Haupt. Effect of the human papillomavirus (HPV) quadrivalent vaccine in a subgroup of women with cervical and vulvar disease: retrospective pooled analysis of trial data. *British Medical Journal*, 344:e1401, 2012.
- [52] E. A. Joura, A. R. Giuliano, O.-E. Iversen, C. Bouchard, C. Mao, J. Mehlsen, E. D. Moreira, Y. Ngan, L. K. Petersen, E. Lazcano-Ponce, P. Pitisuttithum, J. A. Restrepo, G. Stuart, L. Woelber, Y. C. Yang, J. Cuzick, S. M. Garland, W. Huh, S. K. Kjaer, O. M. Bautista, I. S. Chan, J. Chen, R. Gesser, E. Moeller, M. Ritter, S. Vuocolo, and A. Luxembourg. A 9-Valent HPV Vaccine against Infection and Intraepithelial Neoplasia in Women. *New England Journal of Medicine*, 372(8):711–723, 2 2015.
- [53] J. a. Kahn, D. R. Brown, L. Ding, L. E. Widdice, M. L. Shew, S. Glynn, and D. I. Bernstein. Vaccine-type human papillomavirus and evidence of herd protection after vaccine introduction. *Pediatrics*, 130(2):249–56, 8 2012.
- [54] M. J. Keeling and R. Pejman. Formulating the Deterministic SIR Model. In *Modeling Infectious Diseases in Humans and Animals*, chapter 2, pages 15–52. Princeton University Press, Princeton, NJ, 2 edition, 2008.
- [55] T. J. Kemp, A. Hildesheim, M. Safaeian, J. G. Dauner, Y. Pan, C. Porras, J. T. Schiller, D. R. Lowy, R. Herrero, and L. A. Pinto. HPV16/18 L1 VLP vaccine induces cross-neutralizing antibodies that may mediate cross-protection. *Vaccine*, 29(11):2011–2014, 2011.
- [56] J. J. Kim and S. J. Goldie. Health and Economic Implications of HPV Vaccination in the United States. *New England Journal of Medicine*, 359(8):821–832, 8 2008.
- [57] J. J. Kim, K. M. Kuntz, N. K. Stout, S. Mahmud, L. L. Villa, E. L. Franco, and S. J. Goldie. Multiparameter Calibration of a Natural History Model of Cervical Cancer. *American Journal of Epidemiology*, 166(2):137–150, 6 2007.
- [58] A. A. King, E. L. Ionides, C. Bretó, S. P. Ellner, M. J. Ferrari, B. Kendall, M. Lavine, D. Nguyen, D. C. Reuman, H. Wearing, and S. N. Wood. pomp: Statistical Inference for Partially Observed Markov Processes, 2017.

- [59] A. A. King, D. Nguyen, and E. L. Ionides. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software*, 69(12):1–43, 2016.
- [60] S. K. Kjaer, A. J. van den Brule, J. E. Bock, P. A. Poll, G. Engholm, M. E. Sherman, J. M. Wallboomers, and C. J. Meijer. Determinants for genital human papillomavirus (HPV) infection in 1000 randomly chosen young Danish women with normal Pap smear: are there different risk profiles for oncogenic and nononcogenic HPV types? *Cancer Epidemiology, Prevention & Biomarkers*, 6(10):799–805, 1997.
- [61] M. Kohli, N. Ferko, A. Martin, E. Franco, D. Jenkins, S. Gallivan, C. Sherlaw-Johnson, and M. Drummond. Estimating the long-term impact of a prophylactic human papillomavirus 16/18 vaccine on the burden of cervical cancer in the UK. *British Journal of Cancer*, 96(1):143–150, 2007.
- [62] J. Lee, V. Taneja, and R. Vassallo. Cigarette smoking and inflammation: cellular and molecular mechanisms. *Journal of Dental Research*, 91(2):142–149, 2012.
- [63] P. Legendre and L. Legendre. *Numerical Ecology*. Elsevier, 2 edition, 1998.
- [64] K. L. Liaw. A prospective study of human papillomavirus (HPV) type 16 DNA detection by polymerase chain reaction and its association with acquisition and persistence of other HPV types. *Journal of Infectious Diseases*, 183(1):8–15, 2001.
- [65] M. Lipsitch, C. Colijn, T. Cohen, W. P. Hanage, and C. Fraser. No coexistence for free: neutral null models for multistrain pathogens. *Epidemics*, 1(1):2–13, 3 2009.
- [66] M. Lipsitch and J. J. O’Hagan. Patterns of antigenic diversity and the mechanisms that maintain them. *Journal of The Royal Society Interface*, 4(16):787–802, 2007.
- [67] B. Lu, R. P. Viscidi, Y. Wu, J.-H. Lee, A. G. Nyitray, L. L. Villa, E. Lazcano-Ponce, R. J. C. da Silva, M. L. Baggio, M. Quiterio, J. Salmeron, D. C. Smith, M. E. Abrahamsen, M. R. Papenfuss, H. G. Stockwell, and A. R. Giuliano. Prevalent Serum Antibody Is Not a Marker of Immune Protection against Acquisition of Oncogenic HPV16 in Men. *Cancer Research*, 72(3):676–685, 2012.
- [68] B. Lu, Y. Wu, C. Nielson, R. Flores, M. Abrahamsen, M. Papenfuss, R. Harris, and A. Giuliano. Factors Associated with Acquisition and Clearance of Human Papillomavirus Infection in a Cohort of US Men: A Prospective Study. *Journal of Infectious Diseases*, 199(3):362–371, 2009.
- [69] R. Luttmer, M. G. Dijkstra, P. J. F. Snijders, E. S. Jordanova, A. J. King, D. T. M. Pronk, C. Foresta, A. Garolla, P. G. A. Hompes, J. Berkhof, M. C. G. Bleeker, J. Doorbar, D. A. M. Heideman, and C. J. L. M. Meijer. Presence of human papillomavirus in semen of healthy men is firmly associated with HPV infections of the penile epithelium. *Fertility and Sterility*, 104(4):838–844, 2015.
- [70] Y. Ma, R. Madupu, U. Karaoz, C. W. Nossa, L. Yang, S. Yooseph, P. S. Yachinski, E. L. Brodie, K. E. Nelson, and Z. Pei. Human papillomavirus community in healthy persons, defined by metagenomics analysis of human microbiome project shotgun sequencing data sets. *Journal of Virology*, 88(9):4786–97, 5 2014.
- [71] M. J. Mackinnon and K. Marsh. The Selection Landscape of Malaria Parasites. *Science*, 328(5980):866–871, 2010.
- [72] G. A. Maglennon, P. McIntosh, and J. Doorbar. Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology*, 414(2):153–163, 6 2011.
- [73] L. Markowitz, M. Sternberg, E. Dunne, G. McQuillan, and E. R. Unger. Seroprevalence of Human Papillomavirus Types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003–2004. *Journal of Infectious Diseases*, 200(7):1059–1067, 10 2009.
- [74] L. E. Markowitz, S. Hariri, C. Lin, E. F. Dunne, M. Steinau, G. McQuillan, and E. R. Unger. Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003–2010. *Journal of Infectious Diseases*, 208(3):385–393, 2013.
- [75] L. E. Markowitz, G. Liu, S. Hariri, M. Steinau, E. F. Dunne, and E. R. Unger. Prevalence of HPV After Introduction of the Vaccination Program in the United States. *Pediatrics*, 137(3):e20151968, 2016.

-
- [76] S. M. Matthijssse, J. Van Rosmalen, J. A. C. Hontelez, R. Bakker, I. M. C. M. De Kok, M. Van Ballegooijen, and S. J. De Vlas. The role of acquired immunity in the spread of human papillomavirus (HPV): Explorations with a microsimulation model. *PLoS ONE*, 10(2):1–14, 2015.
- [77] J. Monsonego, G. Pollini, M. J. Evrard, P. Sednaoui, L. Monfort, D. Quinzat, R. Dachez, and K. Syrjanen. Linear array genotyping and hybrid capture II assay in detecting human papillomavirus genotypes in women referred for colposcopy due to abnormal Papanicolaou smear. *International Journal of STD & AIDS*, 19(6):385–392, 6 2008.
- [78] A. Moscicki, M. Schiffman, A. Burchell, and G. Albero. Updating the natural history of human papillomavirus and anogenital cancers. *Vaccine*, 30(Supplement 5):F24–F33, 2012.
- [79] A.-B. Moscicki, J. Palefsky, J. Gonzales, and G. K. Schoolnik. Human Papillomavirus Infection in Sexually Active Adolescent Females: Prevalence and Risk Factors. *Pediatric Research*, 28(5):507–513, 1990.
- [80] C. L. Murall, C. T. Bauch, and T. Day. Could the human papillomavirus vaccines drive virulence evolution? *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1798):20141069, 1 2015.
- [81] C. L. Murall, K. S. McCann, and C. T. Bauch. Revising ecological assumptions about Human papillomavirus interactions and type replacement. *Journal of Theoretical Biology*, 350:98–109, 2014.
- [82] M. Nakagawa, D. Stites, S. Patel, S. Farhat, M. Scott, N. Hills, J. Palefsky, and A. Moscicki. Persistence of Human Papillomavirus Type 16 Infection Is Associated with Lack of Cytotoxic T Lymphocyte Response to the E6 Antigens. *Journal of Infectious Diseases*, 182(2):595–598, 8 2000.
- [83] A. Nyitray, R. Carvalho, M. Baggio, L. Beibei, M. Abrahamsen, M. Papenfuss, L. Villa, E. Lazcano-Ponce, and A. Giuliano. Age-Specific Prevalence of and Risk Factors for Anal Human Papillomavirus (HPV) among Men Who Have Sex with Women and Men Who Have Sex with Men : The HPV in Men (HIM) Study. *Journal of Infectious Diseases*, 203(1):49–57, 2011.
- [84] A. G. Nyitray, R. J. Carvalho Da Silva, M. L. Baggio, D. Smith, M. Abrahamsen, M. Papenfuss, H. Y. Lin, M. Quiterio, J. Salmerón, E. Lazcano-Ponce, L. L. Villa, and A. R. Giuliano. Six-month incidence, persistence, and factors associated with persistence of anal human papillomavirus in men: The HPV in men study. *Journal of Infectious Diseases*, 204(11):1711–1722, 2011.
- [85] A. G. Nyitray, M. Chang, L. L. Villa, R. J. Carvalho da Silva, M. L. Baggio, M. Abrahamsen, M. Papenfuss, M. Quiterio, J. Salmerón, E. Lazcano-Ponce, and A. R. Giuliano. The Natural History of Genital Human Papillomavirus Among HIV-Negative Men Having Sex With Men and Men Having Sex With Women. *Journal of Infectious Diseases*, 212(2):202–212, 7 2015.
- [86] A. G. Nyitray, J. Kim, C. H. Hsu, M. Papenfuss, L. Villa, E. Lazcano-Ponce, and A. R. Giuliano. Test-retest reliability of a sexual behavior interview for men residing in Brazil, Mexico, and the United States. *American Journal of Epidemiology*, 170(8):965–974, 2009.
- [87] S.-E. Olsson, S. K. Kjaer, K. Sigurdsson, O.-E. Iversen, M. Hernandez-Avila, C. M. Wheeler, G. Perez, D. R. Brown, L. A. Koutsky, E. H. Tay, P. García, K. A. Ault, S. M. Garland, S. Leodolter, G. W. K. Tang, D. G. Ferris, J. Paavonen, M. Lehtinen, M. Steben, F. X. Bosch, J. Dillner, E. A. Joura, S. Majewski, N. Muñoz, E. R. Myers, L. L. Villa, F. J. Taddeo, C. Roberts, A. Tadesse, J. Bryan, R. Maansson, S. Vuocolo, T. M. Hesley, A. Saah, E. Barr, and R. M. Haupt. Evaluation of quadrivalent HPV 6/11/16/18 vaccine efficacy against cervical and anogenital disease in subjects with serological evidence of prior vaccine type HPV infection. *Human vaccines*, 5(10):696–704, 10 2009.
- [88] U. Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460, 12 1979.
- [89] P. A. Orlando, R. A. Gatenby, A. R. Giuliano, and J. S. Brown. Evolutionary ecology of human papillomavirus: Trade-offs, coexistence, and origins of high-risk and low-risk types. *Journal of Infectious Diseases*, 205(2):272–279, 2012.
- [90] S. J. Pamnani, S. L. Sudenga, R. Viscidi, D. E. Rollison, B. N. Torres, D. J. Ingles, M. Abrahamsen, L. L. Villa, E. Lazcano-Ponce, J. Salmeron, M. Quiterio, Y. Huang, A. Borenstein, and A. R. Giuliano. Impact of Serum Antibodies to HPV Serotypes 6, 11, 16, and 18 to Risks of Subsequent Genital HPV Infections in Men: The HIM Study. *Cancer Research*, 76(20):6066–6075, 10 2016.

- [91] J. M. Partridge, J. P. Hughes, Q. Feng, R. L. Winer, B. A. Weaver, L.-F. Xi, M. E. Stern, S.-K. Lee, S. F. O'Reilly, S. E. Hawes, N. B. Kiviat, and L. A. Koutsky. Genital human papillomavirus infection in men: incidence and risk factors in a cohort of university students. *Journal of Infectious Diseases*, 196(8):1128–1136, 10 2007.
- [92] J.-a. S. Passmore and A.-I. Williamson. Host Immune Responses Associated with Clearance or Persistence of Human Papillomavirus Infections. *Current Obstetrics and Gynecology Reports*, 5(3):177–188, 2016.
- [93] K. U. Petry, D. Scheffel, U. Bode, T. Gabrysiak, H. Köchel, E. Kupsch, M. Glaubitz, S. Niesert, H. Kühnle, and I. Schedel. Cellular immunodeficiency enhances the progression of human papillomavirus-associated cervical lesions. *International Journal of Cancer*, 57(6):836–840, 6 1994.
- [94] C. M. Pierce Campbell, H.-Y. Lin, W. Fulp, M. R. Papenfuss, J. J. Salmerón, M. M. Quiterio, E. Lazcano-Ponce, L. L. Villa, and A. R. Giuliano. Consistent condom use reduces the genital human papillomavirus burden among high-risk men: the HPV infection in men study. *Journal of Infectious Diseases*, 208(3):373–84, 8 2013.
- [95] S. J. Piersma, E. S. Jordanova, M. I. van Poelgeest, K. M. Kwappenberg, J. M. van der Hulst, J. W. Drijfhout, C. J. Melief, G. G. Kenter, G. J. Fleuren, R. Offringa, and S. H. van der Burg. High Number of Intraepithelial CD8+ Tumor-Infiltrating Lymphocytes Is Associated with the Absence of Lymph Node Metastases in Patients with Large Early-Stage Cervical Cancer. *Cancer Research*, 67(1):354–361, 1 2007.
- [96] E. M. Poolman, E. H. Elbasha, and A. P. Galvani. Vaccination and the evolutionary ecology of human papillomavirus. *Vaccine*, 26(Supplement 3):25–30, 2008.
- [97] E. O. Romero-Severson, E. Volz, J. S. Koopman, T. Leitner, and E. L. Ionides. Dynamic Variation in Sexual Contact Rates in a Cohort of HIV-Negative Gay Men. *American Journal of Epidemiology*, 182(3):255–262, 8 2015.
- [98] A. F. Rositch, M. G. Hudgens, D. M. Backes, S. Moses, K. Agot, E. Nyagaya, P. J. F. Snijders, C. J. L. M. Meijer, R. C. Bailey, and J. S. Smith. Vaccine-relevant human papillomavirus (HPV) infections and future acquisition of high-risk HPV types in men. *Journal of Infectious Diseases*, 206(5):669–677, 2012.
- [99] M. Rousseau, J. Pereira, J. Prado, L. Villa, T. Rohan, and E. Franco. Cervical Coinfection with Human Papillomavirus (HPV) Types as a Predictor of Acquisition and Persistence of HPV Infection. *Journal of Infectious Diseases*, 184(12):1508–1517, 12 2001.
- [100] M. D. Ryser, P. E. Gravitt, and E. R. Myers. Mechanistic Mathematical Models: An Underused Platform for HPV Research. *Papillomavirus Research*, 3:46–49, 2017.
- [101] M. D. Ryser, E. R. Myers, and R. Durrett. HPV Clearance and the Neglected Role of Stochasticity. *PLOS Computational Biology*, 11(3):e1004113, 2015.
- [102] M. Safaeian, C. Porras, M. Schiffman, A. C. Rodriguez, S. Wacholder, P. Gonzalez, W. Quint, L.-J. van Doorn, M. E. Sherman, V. Xhenseval, R. Herrero, A. Hildesheim, and f. t. C. R. V. T. Costa Rican Vaccine Trial Group. Epidemiological study of anti-HPV16/18 seropositivity and subsequent risk of HPV16 and -18 infections. *Journal of the National Cancer Institute*, 102(21):1653–62, 11 2010.
- [103] M. Scott, M. Nakagawa, and A. B. Moscicki. Cell-mediated immune response to human papillomavirus infection. *Clinical and Diagnostic Laboratory Immunology*, 8(2):209–20, 3 2001.
- [104] M. L. Shew, A. C. Ermel, Y. Tong, W. Tu, B. Qadadri, and D. R. Brown. Episodic detection of human papillomavirus within a longitudinal cohort of young women. *Journal of Medical Virology*, 87(12):2122–2129, 12 2015.
- [105] M. A. Smith, J.-B. Lew, R. J. Walker, J. M. Brotherton, C. Nickson, and K. Canfell. The predicted impact of HPV vaccination on male infections and male HPV-related cancers in Australia. *Vaccine*, 29(48):9112–9122, 11 2011.
- [106] C. Sonnex, S. Strauss, and J. J. Gray. Detection of human papillomavirus DNA on the fingers of patients with genital warts. *Sexually Transmitted Infections*, 75(5):317–9, 10 1999.
- [107] M. Stanley. Immune responses to human papillomavirus. *Vaccine*, 24(Supplement 1):16–22, 2006.

- [108] M. Stanley, L. A. Pinto, and C. Trimble. Human Papillomavirus Vaccines - Immune Responses. *Vaccine*, 30(Supplement 5):F83–F87, 11 2012.
- [109] M. Steinau, J. M. Onyekwulije, M. Z. Scarbrough, E. R. Unger, J. Dillner, and T. Zhou. Performance of commercial reverse line blot assays for human papillomavirus genotyping. *Journal of Clinical Microbiology*, 50(5):1539–44, 5 2012.
- [110] H. D. Strickler, R. D. Burk, M. Fazzari, K. Anastos, H. Minkoff, L. S. Massad, C. Hall, M. Bacon, A. M. Levine, D. H. Watts, M. J. Silverberg, X. Xue, N. F. Schlecht, S. Melnick, and J. M. Palefsky. Natural History and Possible Reactivation of Human Papillomavirus in Human Immunodeficiency Virus-Positive Women. *Journal of the National Cancer Institute*, 97(8):577–586, 4 2005.
- [111] E. I. Svare, S. K. Kjaer, A. M. Worm, A. Østerlind, C. J. L M Meijer, and A. J. C van den Brule. Risk factors for genital HPV DNA in men resemble those found in women: a study of male attendees at a Danish STD clinic. *Sexually Transmitted Infections*, 78(3):215–218, 2002.
- [112] B. S. Taylor, M. E. Sobieszczyk, F. E. McCutchan, and S. M. Hammer. The challenge of HIV-1 subtype diversity. *New England Journal of Medicine*, 358(15):1590–602, 4 2008.
- [113] R. N. Theiler, S. L. Farr, J. M. Karon, P. Paramsothy, R. Viscidi, A. Duerr, S. Cu-Uvin, J. Sobel, K. Shah, R. S. Klein, and D. J. Jamieson. High-Risk Human Papillomavirus Reactivation in Human Immunodeficiency Virus-Infected Women. *Obstetrics & Gynecology*, 115(6):1150–1158, 6 2010.
- [114] K. Thomas, J. Hughes, J. Kuypers, N. Kiviat, S. Lee, D. Adam, and L. Koutsky. Concurrent and Sequential Acquisition of Different Genital Human Papillomavirus Types. *Journal of Infectious Diseases*, 182(4):1097–1102, 10 2000.
- [115] R. W. Todd, S. Roberts, C. H. Mann, D. M. Luesley, P. H. Gallimore, and J. C. Steele. Human papillomavirus (HPV) type 16-specific CD8+ T cell responses in women with high grade vulvar intraepithelial neoplasia. *International Journal of Cancer*, 108(6):857–862, 2004.
- [116] N. Van de Velde, M. Brisson, and M.-C. Boily. Modeling human papillomavirus vaccine effectiveness: quantifying the impact of parameter uncertainty. *American Journal of Epidemiology*, 165(7):762–75, 2007.
- [117] N. Van de Velde, M. Brisson, and M. C. Boily. Understanding differences in predictions of HPV vaccine effectiveness: A comparative model-based analysis. *Vaccine*, 28(33):5473–5484, 2010.
- [118] J. M. Walboomers. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology*, 189(1):12–19, 1999.
- [119] H. Weinstock, S. Berman, and W. Cates. Sexually Transmitted Diseases Among American Youth: Incidence and Prevalence Estimates, 2000. *Perspectives on Sexual and Reproductive Health*, 36(1):6–10, 1 2004.
- [120] N. Wentzensen, A. C. Rodriguez, R. Viscidi, R. Herrero, A. Hildesheim, A. Ghosh, J. Morales, S. Wacholder, D. Guillen, M. Alfaro, M. Safaeian, R. D. Burk, and M. Schiffman. A Competitive Serological Assay Shows Naturally Acquired Immunity to Human Papillomavirus Infections in the Guanacaste Natural History Study. *Journal of Infectious Diseases*, 204(1):94–102, 7 2011.
- [121] C. M. Wheeler, X. Castellsagué, S. M. Garland, A. Szarewski, J. Paavonen, P. Naud, J. Salmerón, S. N. Chow, D. Apter, H. Kitchener, J. C. Teixeira, S. R. Skinner, U. Jaisamrarn, G. Limson, B. Romanowski, F. Y. Aoki, T. F. Schwarz, W. A. J. Poppe, F. X. Bosch, D. M. Harper, W. Huh, K. Hardt, T. Zahaf, D. Descamps, F. Struyf, G. Dubin, and M. Lehtinen. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncology*, 13(1):100–110, 2012.
- [122] R. L. Winer, S.-K. Lee, J. P. Hughes, D. E. Adam, N. B. Kiviat, and L. A. Koutsky. Genital Human Papillomavirus Infection: Incidence and Risk Factors in a Cohort of Female University Students. *American Journal of Epidemiology*, 157(3):218–226, 2 2003.

Supporting Information

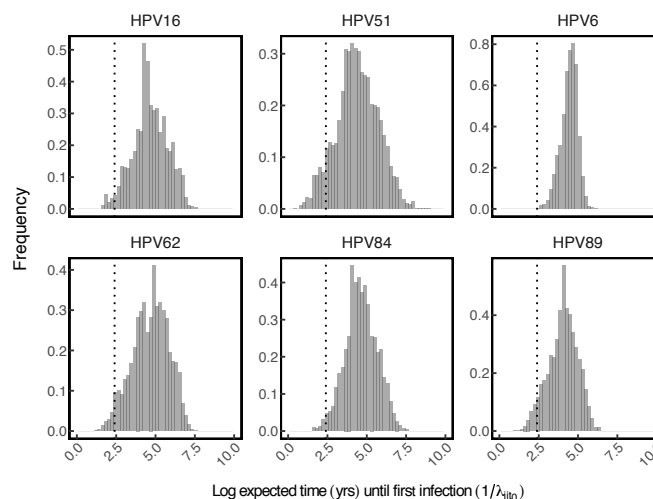


Figure S1: Distribution of the (log) baseline expected time to infection ($t_{\text{expected}} = 1/\lambda_{ij}$) in uninfected individuals assuming no prior infection, such that $\lambda_{ij} = \lambda_{0j} f(\vec{\alpha}_j^T C_{i0})$. For each individual, λ_{ij} was calculated using the maximum likelihood estimate for each element in $\vec{\alpha}$ and the individual-specific covariates C_{i0} , which were reported at the baseline visit ($t = 0$). The y axis reports frequency, while the vertical dashed line in each panel marks an expected time to infection of 10 years. Thus, only the portion of the distribution to the left of the dashed line in each panel represents individuals for whom the expected infection time falls within the next ten years.

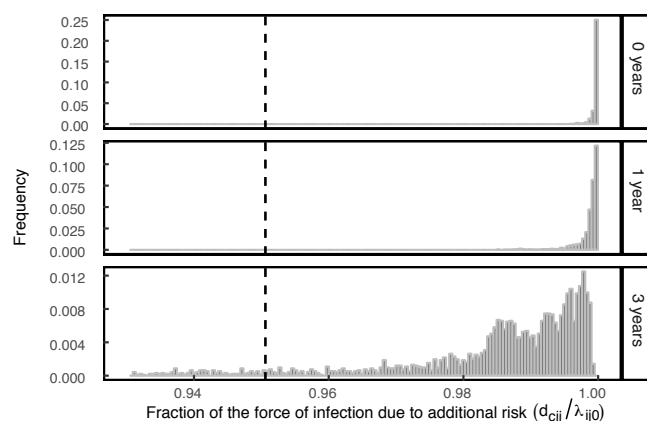


Figure S2: Contribution of the effect of previous infection to overall infection risk in individuals with respect to HPV16. Histograms show the distribution of the fraction of the overall force of infection λ_{ijt} made up by the additional risk d in the population at various times post-clearance of the precedent infection. The mean value for the population-level distribution at each time point is given by the vertical dotted line.

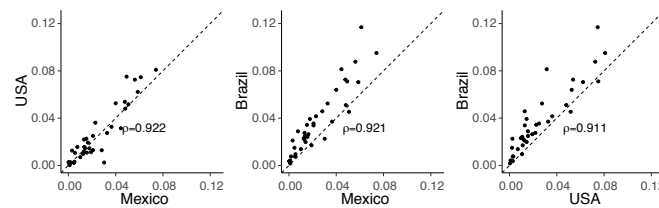


Figure S3: Comparison of the time-averaged prevalence of HPV types in the HIM Dataset. The black dotted lines indicates 1:1 correlation, and p denotes Pearson's correlation coefficient ($p < 10^{-15}$ with 35 degrees of freedom for each comparison).

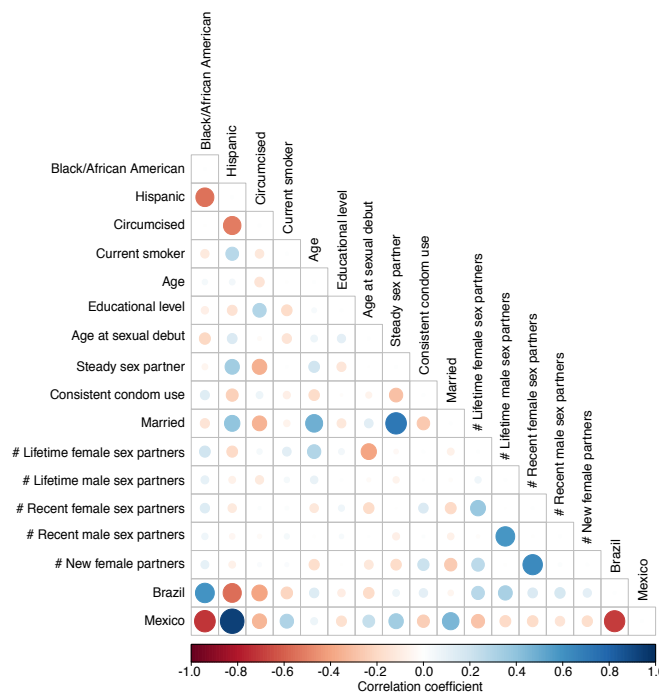


Figure S4: Correlations between the full candidate set of self-reported covariates at the baseline visit. Pearson product-moment correlations were calculated between continuous variables, polyserial correlations (inferred latent correlations between continuous and categorical variables) were calculated between numeric and binary variables, and polychoric correlations (inferred latent correlations between categorical variables) were calculated between binary variables [22, 88]. All correlations shown were significant at the $\alpha = .05$ significance level based on tests for bivariate normality.

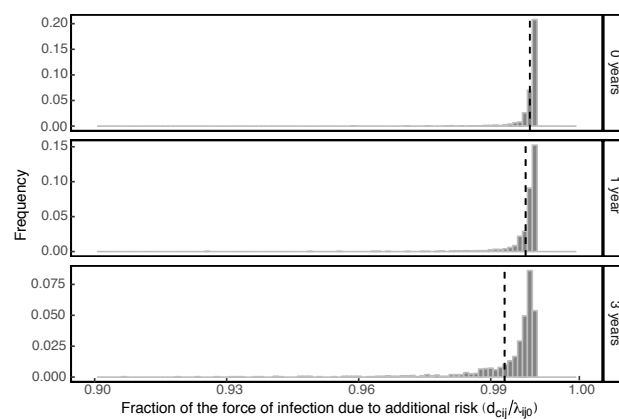


Figure S5: Contribution of the effect of previous infection to the overall infection risk among individuals for HPV16, using a model with 17 covariates. Histograms show the distribution of the fraction of the overall force of infection λ_{ijt} made up by the additional risk d in the population at various times post-clearance of the precedent infection. The mean value for each distribution is given by the vertical dotted line.

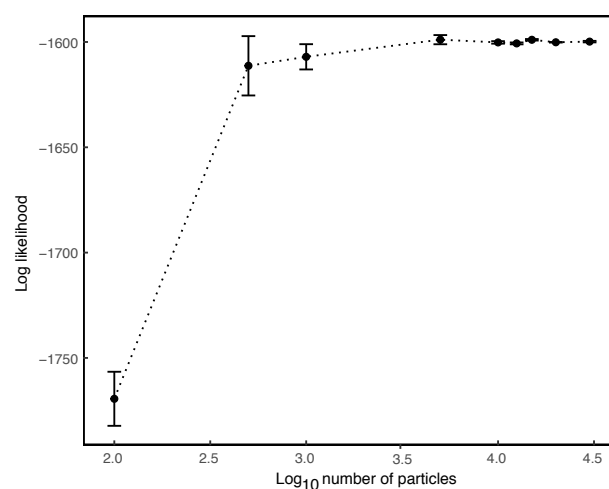


Figure S6: Likelihood of the maximum likelihood parameter set for HPV16 calculated at increasing particle sizes. Point estimates and error bars represent the mean and standard error, respectively, of 10 particle filter replicates.

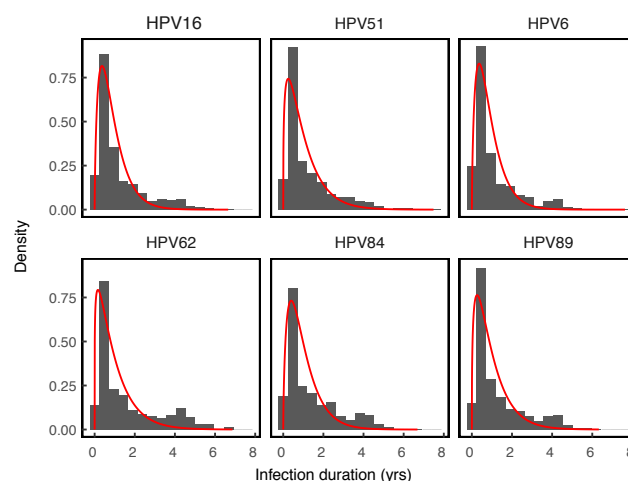


Figure S7: Empirical distribution of infection durations for each HPV type (grey histogram) with an overlying gamma distribution having the same mean and coefficient of variation as in the data (red).

1 Details of the HPV in Men data and selection of covariate variables

We identified 3,656 eligible participants from the 4,123 men enrolled in the HIM study as of October 2014. The eligibility criteria, taken from Giuliano et al. [34], included no prior diagnosis of genital cancer, warts, HIV, or other STIs. Of the eligible participants, we excluded $n = 575$ participants that had missing data at the time of enrollment for the 11 covariates that we included in our three candidate models. We then divided the remaining 3,081 participants into three sexual subclasses based on the number of recent sex partners. For all covariates, "recent" activity indicates activity in the past six months, as reported at each clinic visit. The three sexual subclasses were individuals reporting no recent sexual activity and no recent sex partners (celibate individuals), individuals with one recent sex partner, and individuals with two or more recent sex partners. To account for the effects of sexual subclass on the infection risk, we restricted our analysis to include only the $n = 1,099$ individuals who remained in the same subclass for at least three years over the course of their participation in the study.

Covariates were derived from a risk factor questionnaire that was administered at each visit, which was described and validated previously [34, 86, 69, 91, 23, 1]. Questions covered sociodemographic characteristics, sexual behaviors, sex partnerships, and condom use. Participants were asked to recall their recent behavior, such as recent number of male or female sexual partners, where "recent" referred to behavior over the past six months. Participants had the option of refusing to answer any question, and refusals were treated as missing values as in Giuliano et al. [34], such that a missing covariate was assigned its value at the closest visit.

Covariates were selected for inclusion in the model based on known risk factors identified in the literature for HPV in men [34, 33, 35, 37, 83, 84, 85, 94, 44, 91, 23, 69, 1]. We also included country of residence as a covariate. The full set of candidate covariates (Figure S4) included race (black/African American or other), ethnicity (Hispanic or other), age, age at sexual debut, lifetime numbers of male and female sexual partners, numbers of recent male and female sexual partners, numbers of new male and female sexual partners, presence of a steady sexual partner, marital status, level of education, circumcision status (confirmed by a clinician at the baseline visit), whether or not the participant was a current smoker, whether or not the participant used condoms for the majority of recent sexual encounters, Brazilian nationality, and Mexican nationality, with US nationality used as the baseline. To decrease statistical non-identifiability and to increase computational feasibility, we reduced the full candidate set to a subset of covariates based on observed correlations (Figure S4). Among highly correlated pairs of similar covariates, we heuristically selected one representative. For example, we chose to include recent numbers of sexual partners instead of lifetime numbers of sexual partners and the presence of a steady sexual partner instead of marital status. The variables describing race and ethnicity were strongly correlated with country of origin, so we excluded them. Table S1 gives the final subset of covariates included in the model.

Table S1: Covariates included in the analyses.

	Name	Structure	Notes
1	Age	Continuous	
2	Educational Level	Continuous	Levels: <12 yrs (0), 12 yrs (1), 13-15 yrs (2), 16 yrs (3), >= 17 yrs (4)
3	Age at sexual debut	Continuous	
4	Recent female sexual partners	Continuous	0,1,2,>=3 different partners within the last six months
5	Recent male sexual partners	Continuous	0,1,2,>=3 different partners within the last six months
6	Circumcision status	Binary	
7	Smoking status	Binary	Current smoker
8	Steady sex partner	Binary	Presence of a steady sexual partner
9	Consistent condom use	Binary	Use of a condom more than 50% of the time during intercourse
10	Brazil	Binary	Country of residence
11	Mexico	Binary	Country of residence

2 Alternative and additional models

2.1 Additional risk only model

In this model, the force of infection λ_{ijt} for individual i with type j at time t was determined only by the baseline infection risk for type j and by the effect of previous infection, so that the behavioral and demographic risk factors had no effect. The force of infection is then:

$$\lambda_{ijt} = \lambda_{0j} + I_{\text{prev}} d_{jc_i} e^{-w_j(t-t_{\text{clr}})} \quad (\text{S1})$$

For each HPV type, this model out-performed the memoryless model but it performed much worse than the model that also took into account the covariates (Table S2).

Table S2: Performance of the additional risk only model for each HPV type relative to the full additional risk model.

Type	Log likelihood	SE	ΔAICc
HPV16	-1,636.5	0.6	50.6
HPV51	-1,651.5	1.8	11.7
HPV6	-1,403.6	0.3	8.3
HPV62	-1,877.0	0.4	66.8
HPV84	-1,870.0	0.1	37.3
HPV89	-1,843.4	0.8	22.5

2.2 HPV16/HPV31 interaction model

To test whether infection with HPV31 affects the risk of infection with HPV16, we fitted a model in which the force of infection of HPV16 depends on whether an individual has ever been infected with HPV31. In this model, we included a covariate variable I_{HPV31} , updated at each visit, that indicated whether individual i was currently or previously infected with HPV31. The force of infection λ_{ijt} was thus:

$$\lambda_{ijt} = \lambda_{0j} f(\vec{\alpha}_j \mathbf{C}_{it}) \alpha_{\text{HPV31}} I_{\text{HPV31}} + I_{\text{prev}} d_{jc_i} e^{-w_j(t-t_{\text{clr}})}, \quad (\text{S2})$$

where α_{HPV31} gives the effect of previous or current infection with HPV31 on the risk of infection with HPV16. Our goal was to estimate the direction and strength of interaction between HPV16 and HPV31. Our estimate of the interaction parameter was centered around 1 (MLE 1.3, CI[0.5,2.0]), indicating no significant interaction.

2.3 Additional Covariates model

We added additional covariates to the best-fit model for HPV16, using the full set of 17 original candidate covariates (Figure S4) in the additional risk model. Our goal was to estimate the contribution of the additional risk to the force of infection when allowing for a large number of host risk factors. We used the parameters from the maximum likelihood estimate of the model to estimate the fraction of the force of infection made up by the additional risk d among individuals. This model involved inference for $n=1019$ individuals that had complete covariate information at baseline for the full set of covariates. To compare this model to the best-fit (additional risk) model, we fitted the additional risk model for the same $n=1019$ individuals and obtained the maximum likelihood. We find that the additional risk d still accounts on average for over 90% of the risk for several years after infection clearance (Figure S5). The maximum likelihood and AICc for the model with additional covariates were -1468.4 (SE .63) and 2983.9, respectively, whereas the maximum log likelihood and AICc for the model without additional covariates, fitted to the data from the same individuals, were -1472.8 (SE .86) and 2979.6, respectively. Thus, the ΔAICc for the additional covariates model is 4.3, showing that the improvement in the likelihood from the additional complexity of the additional covariates does not provide a better explanation of the data than the additional risk model.

3 Monte Carlo error in inference from binary panel data

Advances in simulation-based Monte Carlo methods have made it possible to fit complex models to large datasets. We took advantage of extensions of multiple iterated filtering (MIF) [48] to the case of panel data. Iterated filtering uses sequential Monte Carlo, also known as particle filtering, to estimate the likelihood of partially observed Markov process (POMP) models. Sequential Monte Carlo uses stochastic simulations of dynamical models to produce successive populations of weighted particles. Each particle represents a Monte Carlo sample from the probability density of the latent process, conditional on the parameters and the previous observations. As the particle population is propagated along the time series, the particles are weighted and resampled at each data point, and the likelihood of each observation is estimated as the weighted average of the particles.

Large data sets and complex models can result in non-negligible Monte Carlo error in estimated likelihoods. The structure of panel data, a collection of time series that are dynamically independent apart from shared model parameters, yields high Monte Carlo error that often makes it infeasible to calculate the likelihood with an error of less than one log likelihood unit [49]. This is important because a standard approach to calculating 95% confidence intervals relies on the observation that parameter values with log likelihood scores that are within 1.92 units of the maximum log likelihood fall within the 95% confidence interval [9]. Because high levels of Monte Carlo error can make it difficult to accurately estimate likelihoods, high Monte Carlo error rates also make it difficult to estimate 95% confidence intervals. Ionides et al. [49] show that one solution to this problem is to approximate the likelihood in the region of the maximum likelihood by fitting a quadratic to likelihood scores from a large sample of parameter values, which can in turn be used to directly estimate the confidence bounds. Ionides et al. validated this approach with panel data [49], and here we apply it to the case of binary panel data.

To test our approach, before we carried out the model fitting, we quantified the effect of particle size on the likelihood for a given set of parameters. As is often the case in simulation-based approaches, the Monte Carlo error in our simulations is high enough that most particles have very low likelihoods. As a result, our likelihood estimates at first improve rapidly with increases in the number of particles (Figure S6). For particle numbers above about 5000, however, further increases in particle numbers have at most weak effects. We therefore used 20,000 particles for the iterated filtering and 50,000 particles for evaluations of the likelihood by particle filtering. We also accounted for Monte Carlo error in maximum likelihood estimation by initiating a large number of independent MIF searches ($n = 100$) at random parameter values for any given model. Each of the 100 searches began with 200 MIF iterations (.75 cooling factor) and was continued successively (100 iterations, .75 cooling factor) until the maximum likelihood for that particular search was stationary within one log likelihood unit. To identify the MLE for a given model, we required that three searches independently arrive within two log likelihood units of the maximum likelihood value.

4 Model parameters

Table S3: Description of model parameters

Model	Parameter type	Parameter name	Fixed/estimated/nuisance
All Models	Baseline infection risk	λ_{0j}	Estimated
All Models	Infection duration	Shape of gamma distribution (k_j)	Fixed
All Models	Infection duration	Scale of gamma distribution (θ_j)	Fixed
All models	Covariate effects (α)	Age	Estimated
All Models	Covariate effects (α)	Educational status	Estimated
All Models	Covariate effects (α)	Age at sexual debut	Estimated
All Models	Covariate effects (α)	Circumcision status	Estimated
All Models	Covariate effects (α)	Number of recent female sexual partners	Estimated
All Models	Covariate effects (α)	Number of recent male sexual partners	Estimated
All Models	Covariate effects (α)	Steady sexual partner	Estimated
All Models	Covariate effects (α)	Consistent condom use	Estimated
All Models	Covariate effects (α)	Current smoking status	Estimated
All Models	Covariate effects (α)	Mexican nationality (α_{MX})	Estimated
All Models	Covariate effects (α)	Brazilian nationality (α_{BZ})	Estimated
All Models	Initial conditions	Probability of initial infection ($P_{\text{initial},j}$)	Estimated
All Models	Initial conditions	Duration of remaining infection ($F_{\text{remaining}}$)	Nuisance
All Models	Initial conditions	Probability of previous infection (p_{past})	Nuisance
All Models	Measurement model	False positive rate (P_{FP})	Fixed
All Models	Measurement model	False negative rate (P_{FN})	Fixed
Homologous immunity	Immunity dynamics	Magnitude of scaled risk (d)	Estimated
Homologous immunity	Immunity dynamics	Rate of waning of immunity (w)	Estimated
Additional risk model	Additional risk (d)	d_{celibate}	Estimated
Additional risk model	Additional risk (d)	d_{1partner}	Estimated
Additional risk model	Additional risk (d)	d_{multiple}	Estimated

Fixed parameters

The false positive and false negative rates of HPV detection were fixed according to the high sensitivity (96%) and specificity (99%) of the Roche Linear Array HPV genotyping test reported by the manufacturer (Roche Diagnostics), which has been confirmed by other analyses [109, 40, 77]. The duration of each simulated infection with each HPV type j was drawn from a $\Gamma(k_j, \theta_j)$ distribution, where k_j and θ_j were fixed according to the empirical distribution of infection durations in the data for type j (Figure S7). Following Giuliano et al. [34], we required two consecutive negative visits following a positive visit for any HPV type to constitute an observed clearance. Thus, we treated 1-0-1 transitions as false negatives, thereby adjusting data a priori based on the assumption that infection in such cases was actually continuous. Our justification was two-fold: first, this approach allowed us to account for false negatives in HPV sampling beyond the laboratory specifications of the HPV test that were included in the observation model. Second, the approach ensured that the high rates of reinfection estimated by the best-fit model were not the result of failing to account for false clearances, ensuring in turn that our estimates of reinfection rates would be conservative.

Table S4: Values of the parameters describing infection durations

Type	k_j	θ_j
HPV62	1.20	0.81
HPV84	1.67	0.59
HPV89	1.37	0.70
HPV16	1.71	0.52
HPV51	1.33	0.75
HPV6	1.71	0.51

Nuisance parameters

Two nuisance parameters were used to generate the initial conditions for model simulations but were not estimated. For individuals that were initially infected during any simulation, the fraction of an infection duration that they had experienced prior to time $t = 0$ was given by the nuisance parameter $F_{\text{remaining}}$, which was drawn from a uniform (0,1) distribution for each model realization. For individuals that were initially uninfected during any simulation, the probability that the individual had previously been infected at some point in time was given by p_{past} , which was drawn from a uniform (0,1) distribution for each model realization.

Estimated parameters for the best-fit model

Table S5: Values of the estimated parameters. Estimates are provided on a log scale unless otherwise indicated.

Parameter	MLE	Confidence interval	Type
λ_{0j}	-4.6	[-5.2,-4]	HPV62
	-4.6	[-5.2,-4.3]	HPV84
	-4.3	[-4.7,-3.6]	HPV89
	-4.2	[-5.1,-3.5]	HPV16
	-4.3	[-5,-3.8]	HPV51
	-5.1	[-5.7,-4.5]	HPV6
p_{initial} (logit scale)	-2.6	[-2.8,-2.5]	HPV62
	-2.6	[-2.7,-2.5]	HPV84
	-2.8	[-3.1,-2.6]	HPV89
	-2.9	[-3.2,-2.8]	HPV16
	-2.9	[-3.1,-2.8]	HPV51
	-3.2	[-3.4,-2.8]	HPV6
α_{age}	0.0	[-0.3,0.3]	HPV62
	-0.2	[-0.5,0.0]	HPV84
	-0.3	[-0.4,-0.2]	HPV89
	0.0	[-0.2,0.2]	HPV16
	-0.3	[-0.4,0.0]	HPV51
	-0.2	[-0.5,0.0]	HPV6
$\alpha_{\text{Ageatsexualdebut}}$	0.2	[-0.1,0.3]	HPV62
	-0.2	[-0.4,0.1]	HPV84
	-0.3	[-0.4,-0.1]	HPV89
	-0.1	[-0.3,0.1]	HPV16
	-0.3	[-0.5,-0.1]	HPV51
	-0.1	[-0.4,0.1]	HPV6
α_{BZ}	0.9	[0.1,1.2]	HPV62
	0.3	[-0.3,1.0]	HPV84
	0.3	[-0.3,0.7]	HPV89

	0.3	[-0.5,0.6]	HPV16
	-0.3	[-0.5,0.3]	HPV51
	-0.6	[-1.2,0.1]	HPV6
$\alpha_{\text{Circumcised}}$	-0.3	[-0.7,0.1]	HPV62
	0.6	[0.1,0.9]	HPV84
	0.0	[-0.3,0.2]	HPV89
	0.2	[-0.3,0.5]	HPV16
	-0.3	[-0.5,-0.1]	HPV51
	-0.5	[-0.8,0.2]	HPV6
$\alpha_{\text{Consistent condom use}}$	-0.2	[-0.8,0.1]	HPV62
	-0.7	[-1.4,-0.3]	HPV84
	-0.1	[-0.8,0.1]	HPV89
	-0.3	[-0.8,0.1]	HPV16
	-0.2	[-0.5,0.1]	HPV51
	0.3	[-0.1,0.5]	HPV6
$\alpha_{\text{Current smoker}}$	0.5	[0.2,0.8]	HPV62
	0.8	[0.5,1.2]	HPV84
	0.1	[-0.1,0.4]	HPV89
	-0.2	[-0.6,0.3]	HPV16
	0.4	[0.2,0.7]	HPV51
	0.6	[0.2,1.0]	HPV6
$\alpha_{\text{\#Recent female sex partners}}$	0.6	[0.5,0.8]	HPV62
	0.6	[0.5,0.9]	HPV84
	0.6	[0.5,0.8]	HPV89
	0.8	[0.6,0.9]	HPV16
	0.6	[0.4,0.8]	HPV51
	0.6	[0.5,1.8]	HPV6
$\alpha_{\text{\#Recent male sex partners}}$	0.0	[-0.5,0.2]	HPV62
	0.2	[0.0,0.4]	HPV84
	0.5	[0.3,0.7]	HPV89
	-0.8	[-1.9,-0.3]	HPV16
	0.1	[-0.1,0.3]	HPV51
	0.2	[-0.1,0.6]	HPV6
$\alpha_{\text{Educational level}}$	-0.2	[-0.4,0.0]	HPV62
	0.0	[-0.3,0.1]	HPV84
	0.1	[0.0,0.2]	HPV89
	0.0	[-0.2,0.1]	HPV16
	0.0	[-0.2,0.2]	HPV51
	0.0	[-0.2,0.1]	HPV6
α_{MX}	-1.3	[-1.9,-0.6]	HPV62
	-0.1	[-0.8,0.4]	HPV84
	-0.5	[-1.0,0.0]	HPV89
	-1.1	[-1.4,-0.7]	HPV16
	-1.0	[-1.4,-0.7]	HPV51
	-0.4	[-0.8,0.3]	HPV6
$\alpha_{\text{Steady sex partner}}$	-0.7	[-1.0,-0.3]	HPV62
	-0.9	[-1.2,-0.3]	HPV84
	-0.5	[-0.7,-0.3]	HPV89
	-0.9	[-1.2,-0.5]	HPV16
	-0.2	[-0.5,0.2]	HPV51
	-0.4	[-1.0,0.0]	HPV6
d_{celibate}	2.7	[2.2,3.3]	HPV62
	1.0	[0.1,1.6]	HPV84

	5.2	[4.3,6.2]	HPV89
	4.0	[3.5,4.9]	HPV16
	3.9	[2.4,4.2]	HPV51
	2.1	[1.3,2.9]	HPV6
$d_{1 \text{ partner}}$	3.3	[2.9,3.7]	HPV62
	1.5	[-0.3,3]	HPV84
	4.2	[3.5,4.8]	HPV89
	3.8	[3.2,4.7]	HPV16
	3.6	[2.5,4.1]	HPV51
	2.8	[2.2,3.3]	HPV6
$d_{>1 \text{ partner}}$	3.1	[2.7,3.6]	HPV62
	0.7	[0.3,3.4]	HPV84
	3.9	[3.1,4.5]	HPV89
	4.4	[3.8,5.2]	HPV16
	3	[2.7,3.9]	HPV51
	2.3	[1.7,2.9]	HPV6
w	-0.3	[-0.5,-0.1]	HPV62
	-0.5	[-0.8,-0.3]	HPV84
	0	[-0.3,0.2]	HPV89
	0.3	[0.1,0.5]	HPV16
	-1.3	[-1.6,-0.8]	HPV51
	0	[-0.2,0.3]	HPV6

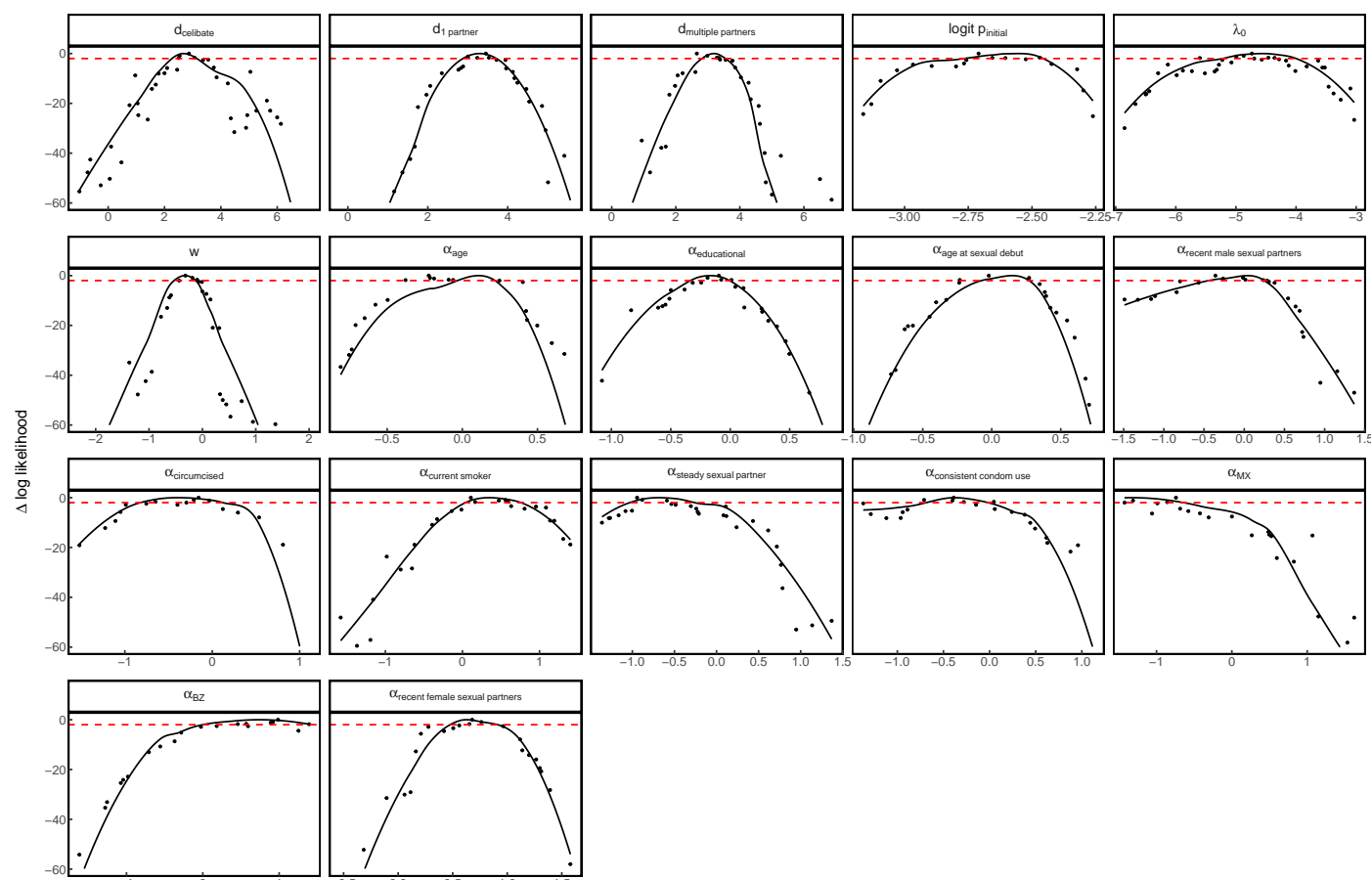


Figure S8: HPV62: Likelihood profiles for each estimated parameter. The x axis gives the log of the profile parameter value unless the logit distribution is specified. The y axis gives the likelihood relative to the maximum likelihood for the additional risk model for HPV62. The red dashed horizontal line indicates the cutoff of 1.92 log likelihood units used to determine the confidence interval.

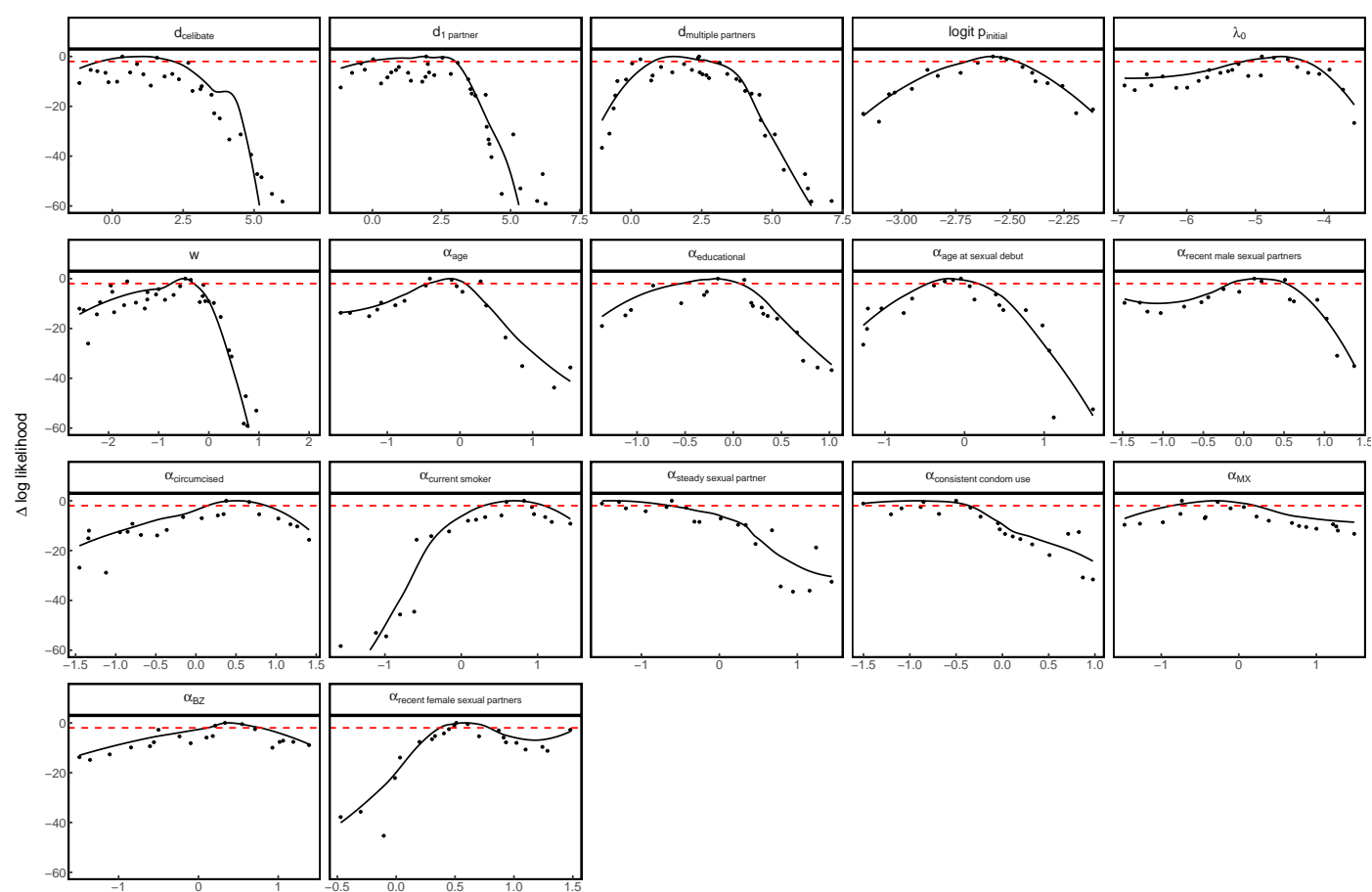


Figure S9: HPV84: Likelihood profiles for each estimated parameter, as in Figure S8.

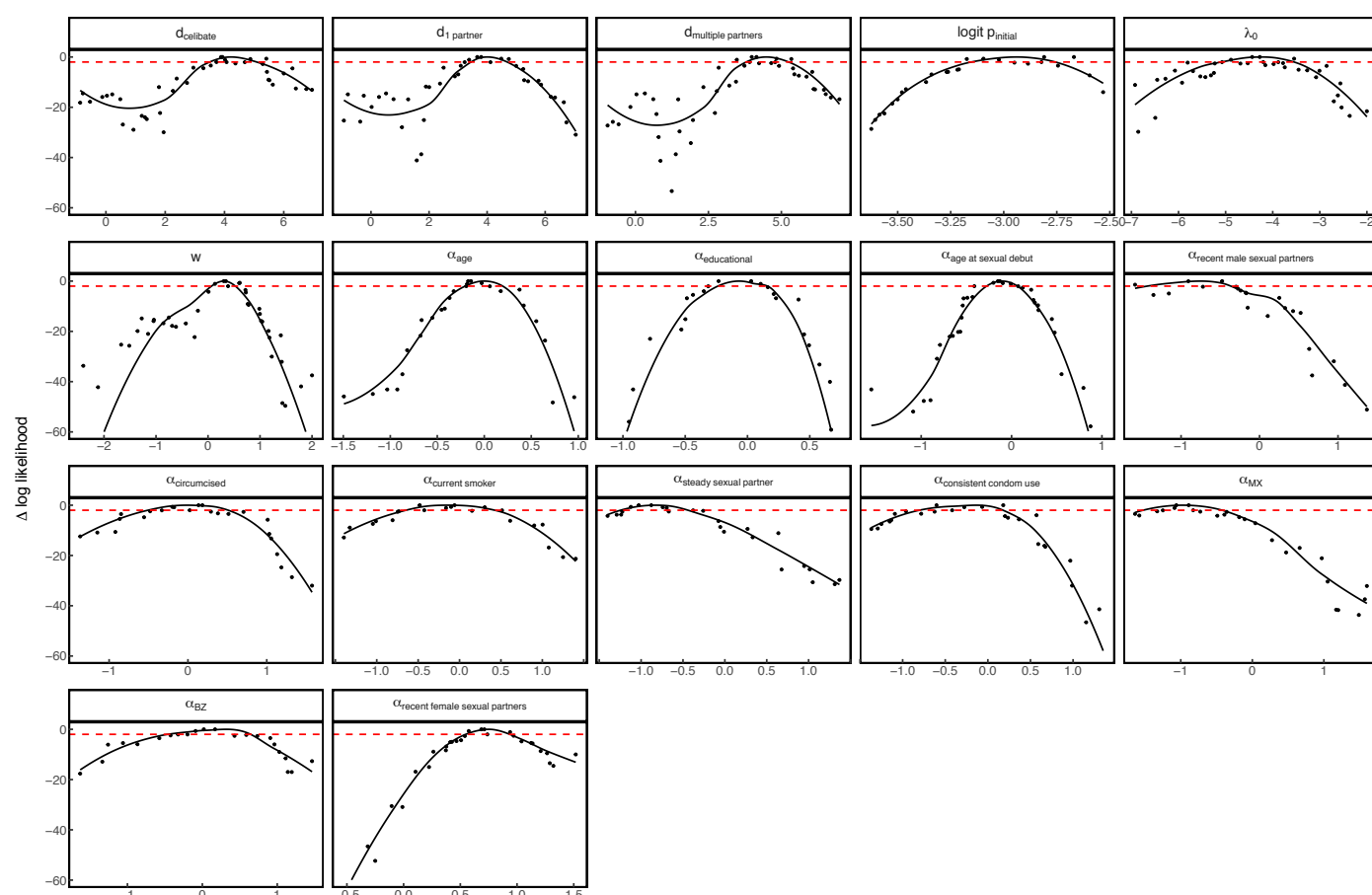


Figure S10: HPV16: Likelihood profiles for each estimated parameter, as in Figure S8.

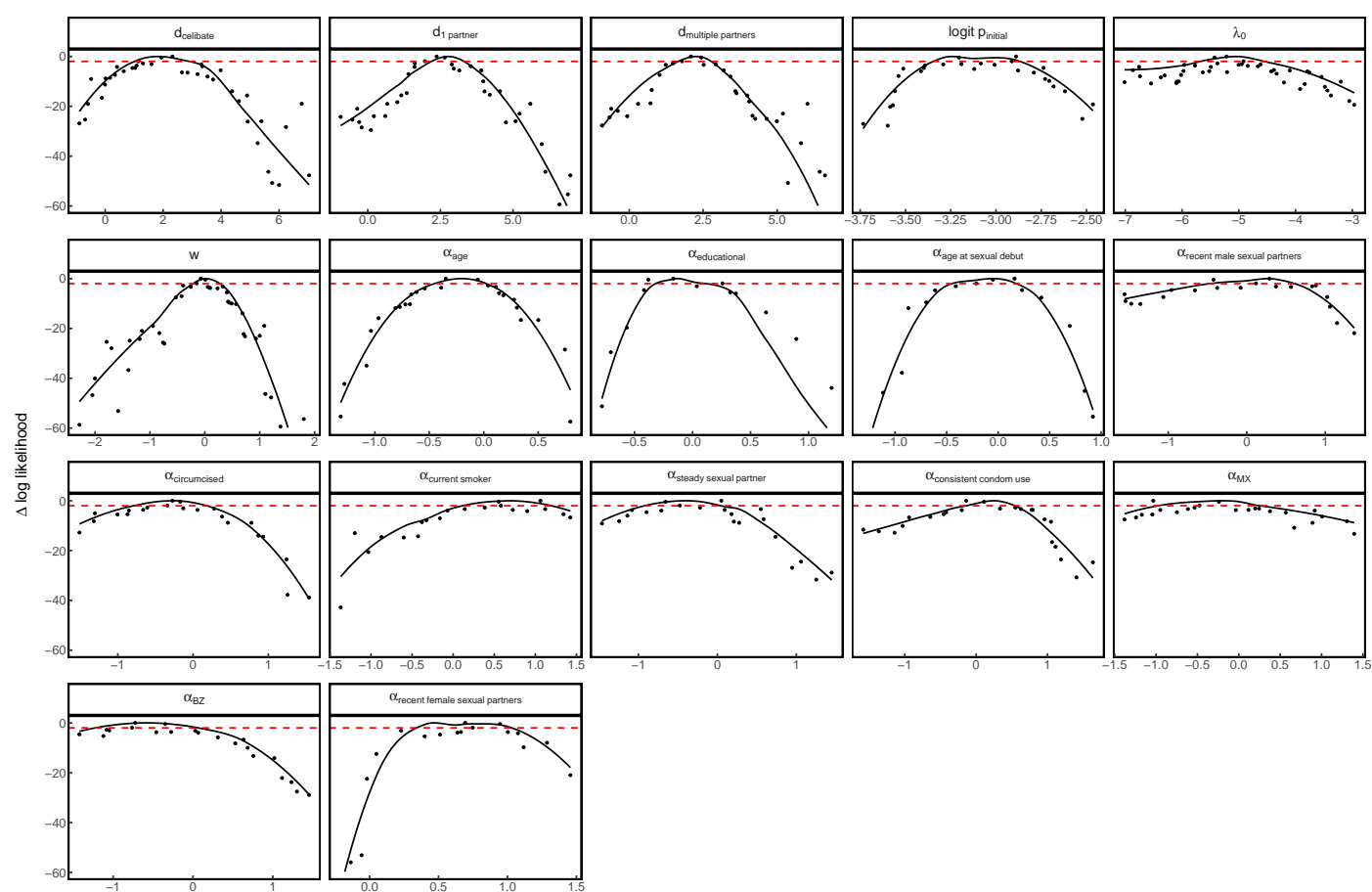


Figure S11: HPV6: Likelihood profiles for each estimated parameter, as in Figure S8.

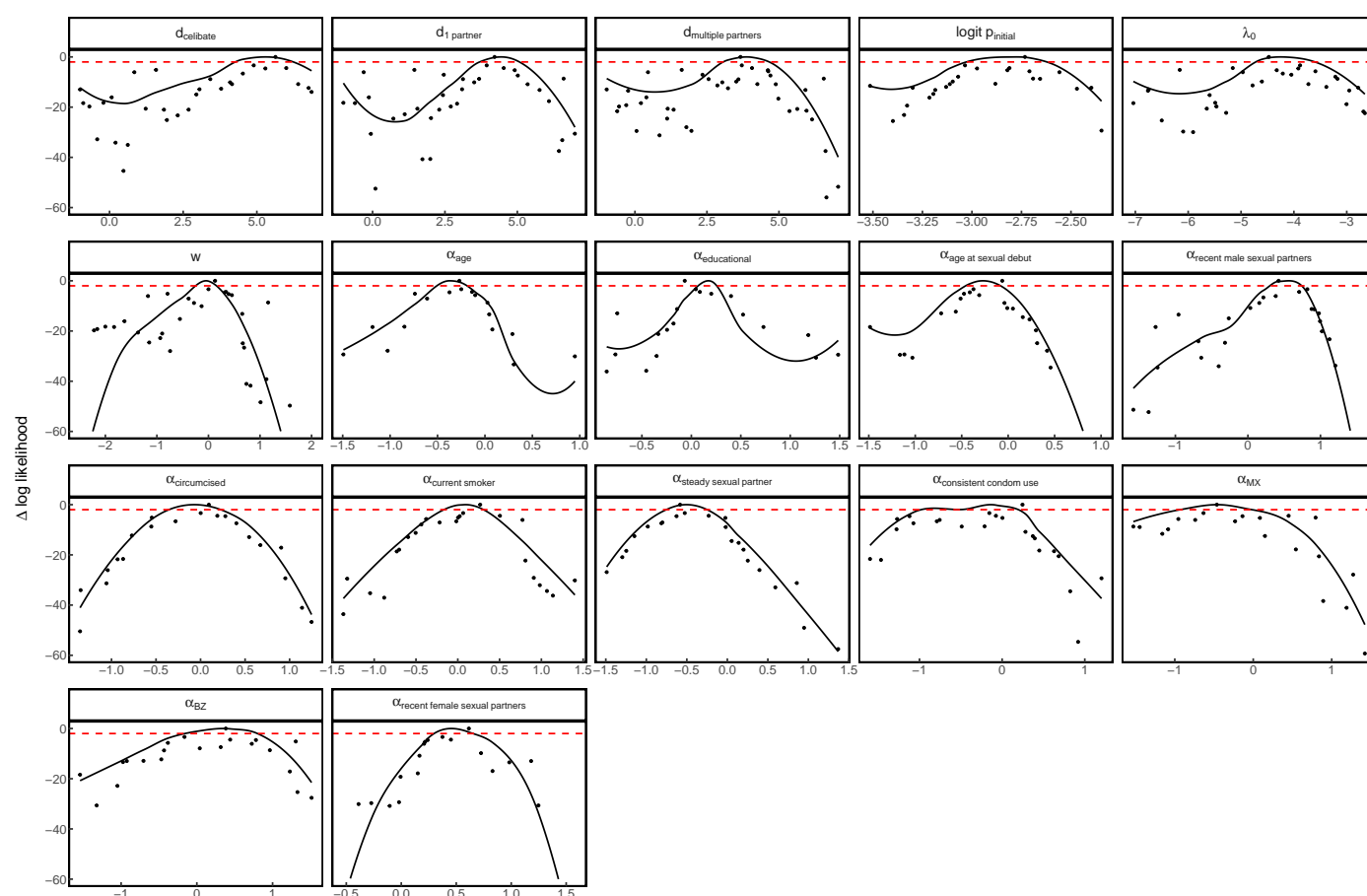


Figure S12: HPV89: Likelihood profiles for each estimated parameter, as in Figure S8.

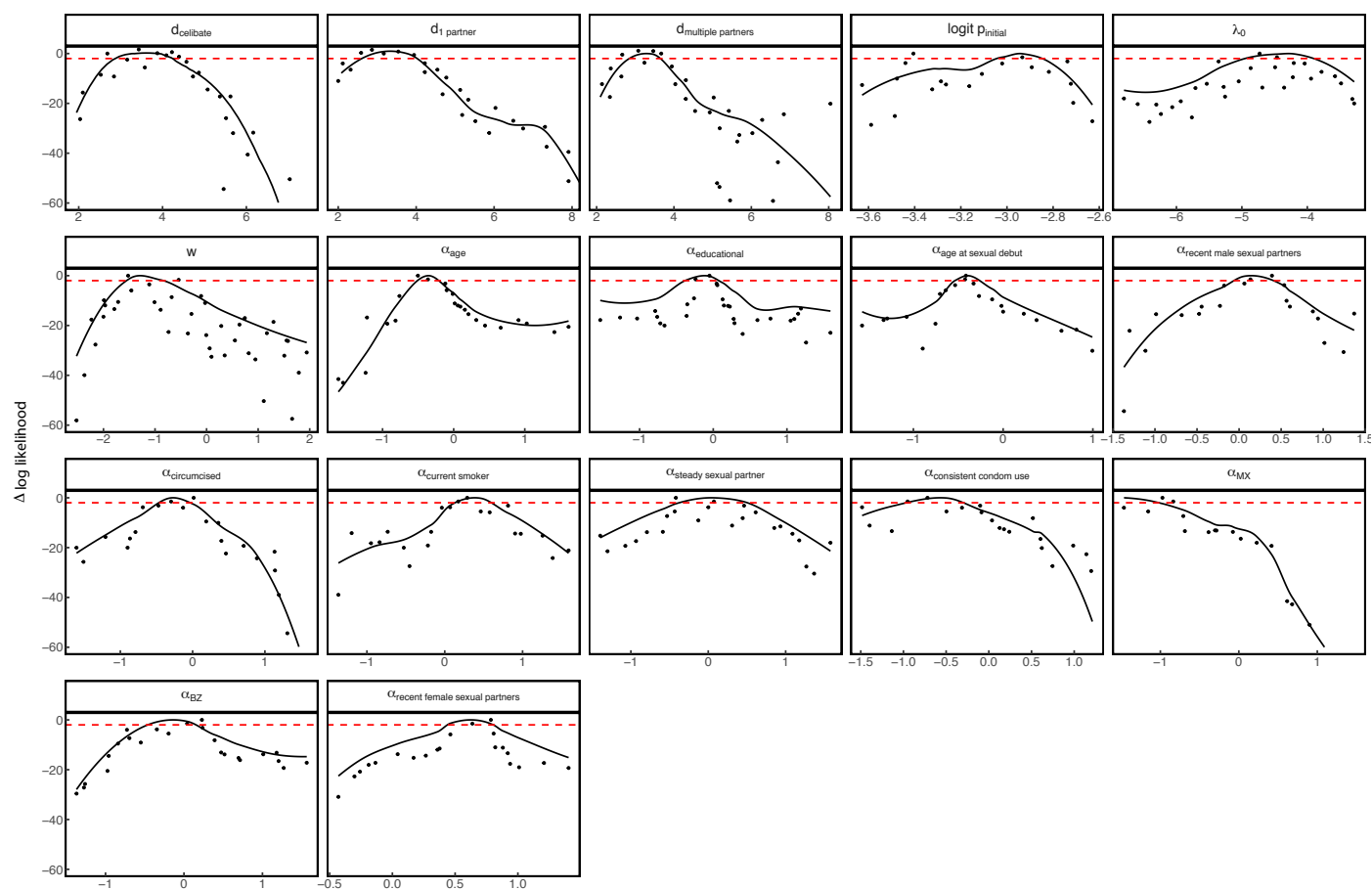


Figure S13: HPV51: Likelihood profiles for each estimated parameter, as in Figure S8.