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 afterward. For HPV16, the type responsible for most HPV-related cancers, an initial infection increases the 1-year probability of reinfection by 20-fold, and the probability of reinfection remains 14-fold higher 2 years later. This increased risk occurs in both sexually active and celibate men, suggesting that it arises from autoinoculation, episodic reactivation of latent virus, or both. Overall, our results suggest 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. Because of the high risk of reinfection, vaccinating boys who have not yet been exposed may be crucial to reduce prevalence, but our results suggest that there may also be large benefits to vaccinating previously infected individuals.

Infectious disease | Mathematical model | Ecology | Public health | Vaccination

Human papillomavirus (HPV), a major cause of genital warts and anogenital and oropharyngeal cancers (1), is the most common sexually transmitted infection (2). While the population prevalence of genital HPV is ~40% among women and 45% among men in the United States, over 200 HPV types have been identified, and the prevalence of individual types never exceeds 10% (3). The quadrivalent vaccine has been effective against the HPV types that cause the most disease (4), and a recent nine-valent vaccine covers additional oncogenic types (5). Predicting the effects of multivalent vaccines is necessary to predict the effects of multivalent vaccines. We investigate HPV transmission dynamics by fitting mechanistic models to data from unvaccinated men. We find that infection risk in a susceptible population is, on average, low and concentrated in high-risk individuals and that, rather than inducing protective immunity, HPV infection strongly increases the risk of future infection by the same type. Thus, high HPV prevalence results from frequent reinfection or persistent infections within individuals. Vaccinating boys before their first HPV exposure may be especially effective in reducing prevalence, and vaccinating previously infected men may also be effective.

Significance

Human papillomavirus (HPV) is the most common sexually transmitted infection. Understanding the factors underlying the prevalence and diversity of over 200 HPV types is necessary to predict the effects of multivalent vaccines. We investigate HPV transmission dynamics by fitting mechanistic models to data from unvaccinated men. We find that infection risk in a susceptible population is, on average, low and concentrated in high-risk individuals and that, rather than inducing protective immunity, HPV infection strongly increases the risk of future infection by the same type. Thus, high HPV prevalence results from frequent reinfection or persistent infections within individuals. Vaccinating boys before their first HPV exposure may be especially effective in reducing prevalence, and vaccinating previously infected men may also be effective.

Author contributions: S.L.R., G.D., and S.C. designed research; S.L.R., L.L.V., E.L.-P., and A.R.G. performed research; E.B.B. and V.D. contributed new reagents/analytic tools; S.L.R. analyzed data; V.D. provided statistical expertise for the mathematical models; L.L.V., E.L.-P., and A.R.G. contributed the clinical data; A.R.G. provided epidemiological expertise to the analysis; and S.L.R., G.D., and S.C. wrote the paper.

Conflict of interest statement: L.L.V. is a consultant and member of the board of Merck, Sharp & Dohme for the HPV prophylactic vaccine. The institution of A.R.G. has received grant funds on her behalf from Merck & Co for investigator-initiated research. A.R.G. also serves on the Scientific and Advisory Boards for Merck & Co. Neither of these relate to the submitted work. All other authors declare no conflict of interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

1To whom correspondence should be addressed. Email: slr@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714712114/-/DCSupplemental.

Edited by Burton H. Singer, University of Florida, Gainesville, FL, and approved November 10, 2017 (received for review August 23, 2017)
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 y (16–18), suggesting protective homologous immunity. In this scenario, the elevated cancer risk associated with particular HPV types results from a small fraction of persistent infections (18). Longitudinal studies, however, have shown that hosts can be infected repeatedly (19). Although type-specific antibodies may provide modest protection against future infection in women (20), serum antibody is not a marker of immune protection in men (21, 22). Thus, the strength of homologous immunity remains unclear.

Evidence for competition between HPV types is also conflicting. Immune-mediated competition has been invoked to explain small increases in the prevalence of nonvaccine HPV types after HPV vaccination (23) and to explain weak cross-protection from the vaccine against related nonvaccine types (24). Mathematical models show that type competition is consistent with observed patterns of HPV prevalence (25, 26). Nevertheless, there is little empirical evidence for intertype competition (27), as shown in part by elevated rates of multiple-type compared with single-type infections and frequent concurrent acquisition of HPV types (28).

Meanwhile, the risk of HPV infection depends on differences in demographic and behavioral risk factors between host sub-populations. For example, a host’s number of sexual partners strongly affects infection risk (1, 18, 29), and some evidence suggests that oncogenic and nononcogenic types are differently sensitive to the numbers of partners among hosts (30). However, detailed comparisons of risk factors for infection with each HPV type have not been performed. Different risk factors would suggest differences between types among host subpopulations that could explain type prevalence.

To investigate the factors determining 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 underlie pathogen dynamics, but HPV has received comparatively little attention (17). HPV models have generally focused on qualitative dynamics (25, 31–33) and on predicting disease for health policy (34, 35), relying on informal methods to estimate parameters. Detailed longitudinal studies present an opportunity to use mechanistic models to rigorously test hypotheses with 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 within and 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 6-mo intervals over 5 y (36). In these data, the mean prevalence of all HPV is 65%, but no type has shown coexistence. The rank prevalence of HPV types is also constant among geographical locations, although absolute prevalence is higher in Brazil than in the United States or Mexico (SI Appendix, Fig. S4).

We analyzed the five HPV types with the highest mean prevalence: HPV62, HPV84, HPV89, HPV16, and HPV51. We also analyzed HPV6, the ninth most prevalent type, because it is included in the quadrivalent vaccine and is highly associated with genital warts (1). We accounted for previously identified risk factors for HPV infection in the HIM Study dataset (29, 36). The risk factors for any HPV type include markers of increased exposure to infected sexual partners as well as nonsexual behaviors, such as tobacco use, and demographic traits. Several factors, notably circumcision and sexual orientation, differ in their effects among HPV types (37, 38). Type-specific differences may reflect meaningful ecological distinctions. We therefore modeled the effects of a diverse set of risk factors.

Past Infection with HPV Confers Minimal Protection Against Infection with the Same Type. Our models test three hypotheses about HPV infection. Under the simplest or “memoryless” model, infection risk depends only on the effects of host- and HPV type-specific risk factors (Eq. 1), with no consideration of immunity. Two more complex models account for the effects of previous infection with the same type (see Materials and Methods for model details).

Our three models differ only in their assumptions about the instantaneous per capita infection risk, or “force of infection” (39), for host i with HPV type j at time t given by λijt.

i) In the memoryless model, the force of infection depends only on host-specific risk factors and the type-specific baseline force of infection, λ0j:

\[ \lambda_{ijt} = \lambda_0, f(\alpha_i C_i) \]  

[1]

The vector \( \alpha_i \) scales the effect of each of the \( M \) covariates on the force of infection, where \( C_i \) is the covariate matrix:

\[ f(\alpha_i C_i) = e^{\alpha_{i1} C_{i1} + \alpha_{i2} C_{i2} + \alpha_{i3} C_{i3} + \ldots + \alpha_{iM} C_{iM}} \]  

[2]

Each model included five continuous or ordinal covariates and six binary covariates (\( M = 11 \) individual-level risk factors).

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

\[ p_{\text{first}} = (1 + \Delta_t) (1 - e^{-\lambda_0, f(\alpha_i C_i) \Delta_t}) \]

\[ p_{\text{subsequent}} = p_{\text{first}} (1 - (1 - d) e^{-w(t - t_{\text{first}})}) \]

The strength of immunity \( d \) is constrained to fall between zero and one, so that \( d \) reduces the probability of reinfection...
over the time interval $\Delta t$. Note that, for $d = 1$, the homologous immunity model is identical to the memoryless model. After the previous infection clears at time $t_{\text{clr}}$, immunity wanes at rate $w$.

iii) In the additional risk model, the risk of an initial infection is determined as in the memoryless model (Eq. 1), but the risk of reinfection is allowed to be higher. The force of infection thus includes an additional risk factor, which is distinct from the host-specific covariate effects (such as the number of recent sexual partners) and describes only the effect of previous infection (Eq. 4 and Fig. 2):

$$\lambda_{ijt} = \lambda_{ij0} f(C_{it}) + \lambda_{ij1} d_{ijt} e^{-w(t - t_{i,\text{clr}})}.$$  \[4\]

The additional risk $d_{ijt}$ depends on the infecting type $j$ and on the sexual subclass $c_{it}$ of host $i$. The variable $\lambda_{ij1}$ indicates whether there was a previous infection with type $j$.

We included an observation model that determines the probability of the observed data using the sensitivity and specificity of the HPV genotyping test. We fitted our models to the data using a likelihood-based nonlinear fitting routine (SI Appendix) (40). For model selection, 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 (41). The best model has the smallest AICc value.

We first tested whether infection with an HPV type reduces the risk of subsequent infection with the same type, allowing for variation in the degree and duration of protective homologous immunity [3]. Notably, the homologous immunity model reduced to the memoryless model for each HPV type ($d = 1$) (Eq. 3) at the point estimate of the maximum likelihood. The likelihood of the homologous immunity model for each type was thus effectively identical to that of the memoryless model (SI Appendix, Table S1). Small discrepancies in the likelihoods in the two models arise from the Monte Carlo error associated with the use of simulation-based estimation of the likelihood (SI Appendix). The additional parameter in the homologous immunity model therefore increased model complexity without meaningfully improving the likelihood, yielding higher AICc scores. There is thus no evidence for homologous immunity against any of the HPV types that we studied, suggesting that intratype competition is weak or absent.

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 can 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 (42, 43). 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 (SI Appendix). Our estimate of the additional risk of HPV16 infection given previous HPV31 infection is centered around 1 (1.3; 95% confidence interval: 0.5, 2.0), suggesting that the two types do not compete. This lack of an effect may be partly caused by the low prevalence of HPV31 (Fig. 1) and correspondingly low statistical power, but the lack of even a trend toward competition nevertheless suggests that there is no interaction. We therefore conclude that neither intratype nor intertype 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 for all six HPV types (SI Appendix, Table S1). Here, we included individuals who were infected at baseline, but excluding such individuals yielded the same result (SI Appendix). To determine whether the additional risk was caused by repeated exposure to the same HPV type by infected sexual partners, we fitted the magnitude of the additional risk $d$ separately for three sexual subclasses: celibate individuals, individuals reporting one recent sexual partner, and individuals reporting multiple recent sexual partners. We included only people who remained in the same sexual subclass for at least 3 consecutive years. Celibate individuals were defined as reporting no recent receptive or insertional anal, vaginal, or oral sex with male or female partners (SI Appendix). For all six types, previous infection significantly raises the risk of subsequent homologous infection in each sexual subclass (Fig. 3A). The high additional risk experienced by celibate individuals ($d_{\text{celibate}}$) 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, $\lambda_{ij0}$ (SI Appendix, Fig. S1), showing that the high estimated additional risk does not reflect statistical nonidentifiability between additional risk and baseline risk. Estimates of the additional risk between different sexual subclasses have the same magnitude (Fig. 3A), showing that variation in the additional risk across sexual subclasses is minimal.

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 $3$ y after clearing the previous infection. Even several years after the initial infection is cleared, the additional risk caused by previous infection accounts for more than 90% of the force of infection (SI Appendix, Fig. S3). Moreover, immediately after infection, the 1-y probability of reinfection with HPV16 (Eq. 8) is, on average, 20.4-fold higher than the estimate of baseline infection risk, $\lambda_{ij0}$ (SI Appendix, Fig. S3).
than the probability of infection in a naïve individual. The average age increase is 19.1- to 20.5-fold among HPV types. Even 3 y after clearing an infection, the probability of reinfection with HPV16 in the following year remains 13.5-fold higher (7.4- to 20.5-fold among types).

The strong additional risk independent of sexual subclass is consistent with two major biological explanations: repeat infections, presumably caused by autoinoculation between anatomical sites, or episodic reactivation of latent infection. Because our models account for 11 different risk factors other than previous infection, it is unlikely that the increased risk is because of confounding by unmeasured individual-level risk factors. Completely ruling out unmeasured covariates is impossible without controlled experiments. As an initial test, however, we repeated our estimation of the additional risk $d$ using a model that includes additional measured covariates (17 in total). In this more complex model, the additional risk $d$ still accounts for more than 90% of the force of infection after 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 (SI Appendix). This result suggests that the increased risk is not simply because of unmeasured individual-level risk factors, 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 (Eq. 4) for all types (SI Appendix). Moreover, the effects of the host-specific risk factors vary among HPV types. To understand 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 ($1/\lambda_{0ij}$), a measure of risk, is generally low but varies by orders of magnitude among individuals in the naïve population (SI Appendix, Fig. S2). Thus, the risk of initial infection with any type is concentrated in a few high-risk individuals. This feature captures a major part of the data. Under the assumption that the additional risk is a constant fraction of the population prevalence of an HPV type arises from repeated infections in relatively few individuals. In the memoryless model, by contrast, the initial infection risk is higher, previous infection has no effect on subsequent risk, and prevalence arises from fewer infections in more individuals. Simulations from the maximum likelihood estimates of the two models show that both reproduce the population-level prevalence in the data across visits (SI Appendix, Fig. S9). However, only the additional risk model accurately predicts the fraction of individuals who ever experience infection, which the memoryless model overestimates (Fig. 3B). Unmeasured assortative mixing (44) and simultaneous exposure to multiple HPV types through sexual partnerships (45) can confound associations between HPV incidence and past exposures. Residual confounding may be a problem, because the initial infection risk is low and sensitive to host-specific risk factors. However, such confounding would be minimized in celibate individuals. We, therefore, reestimated the magnitude of the additional risk $d_{celibate}$ by fitting the additional risk model (Eq. 4) to data from celibate individuals only. In the celibates-only analysis, the risk conferred by previous infection was still high (SI Appendix) and consistent with our previous estimates. The high additional risk inferred for the celibate individuals and the consistency of the additional risk for all three sexual subclasses suggest that the additional risk reflects autoinoculation or reactivation rather than confounding from unmeasured sexual activity.

The effects of each covariate also show interesting similarities and differences between types (Fig. 4). Having more 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 80–120% among types compared with 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 types, indicating subtle differences between types. Our results are largely consistent with previous findings, with differences likely arising from the greater statistical power of our inference technique (SI Appendix).

**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. Our results show little evidence of competition within and between HPV types, and instead, they show high rates of reinfection or persistence by the same type. 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. Our results instead suggest that high HPV prevalence results from continuous or repeated infection with multiple, apparently independent virus types. Types differ slightly in their risk factors, suggesting that ecological niche partitioning could further promote coexistence.

Our results add to growing evidence that HPV infection, in contrast to vaccination, does not confer protective homologous immunity in men. Although we allowed the strength and duration of immunity to vary after initial infection, we could not detect any protection against reinfection with the same type. Because sterilizing immunity reduces infection rates, models that assume that infection induces protective immunity in men (33–35, 46) likely underestimate vaccine effectiveness. Because the genetic background of immunity in men supports the hypothesis that genital infection differs between men and women. Although the durations and distributions of infections in genital HPV infection are similar in men and women.

![Figure 4](https://www.pnas.org/cgi/doi/10.1073/pnas.1714712114)
(1, 3), the prevalence of genital HPV is higher in men (3). Acquired immunity has been proposed to explain declining cervical HPV prevalence by age in women in some countries (47, 48), whereas HPV prevalence does not change with age in men (3, 36). Furthermore, the seroprevalence of some types is higher in women than in men from the same source population (49), although homologous protection in women is still likely to be weak (20) for HPV-related lesions (56). Similarly, skin infections may instead result from increases in susceptibility to autoinoculation.

The high risk of recurring infection is consistent with either autoinoculation across anatomic sites or episodic reactivation of latent infection. Studies showing type-concordant HPV infection across anatomic sites support the importance of autoinoculation. First, HPV DNA has been detected on the fingers of patients with genital warts (50), suggesting transmission between the hands and the genitals. Second, an analysis of a subcohort of the HIM Study showed a 1.5- to 15-fold increase in the risk of anal HPV infection after genital infection (38), and a similar analysis of women found a 20-fold increase in the risk of recurrent anal HPV infection given infection at the cervix (51). Moreover, 63% of the cases of recurrent anal HPV occurred without a self-reported history of anal sex (51). Significantly, the magnitude of the effect in these studies is consistent with our estimate that previous infection yields a 20-fold increase in the risk of subsequent infection.

Apparent reinfection could instead be caused by the reactivation of latent virus. Whether HPV persists latently remains unknown (17, 52), but an animal model of cottontail oral papillomavirus provides evidence of a latent reservoir (53). Furthermore, sexually inactive HIV-positive women have a higher risk of recurrent HPV infection compared with HIV-negative controls (54), and among HIV-positive females, higher CD4+ T-cell counts are negatively correlated with recurrent infection risk (55). The latter two studies have been interpreted as evidence of reactivation of latent infection (52), but either effect may instead result from increases in susceptibility to autoinoculation.

Furthermore, the studies in HIV-positive women suggest that reactivation requires immune suppression and may therefore be rare in healthy individuals. Vaccine efficacy trials 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 (56). Similarly, one study showed 100% efficacy of the HPV vaccine against HPV-related disease in individuals with serological evidence of past HPV6, -11, -16, or -18 infection (57). HPV vaccines prevent infection by inducing antibodies that block viral entry, and such antibodies would likely not prevent reactivation. If indeed vaccines do not affect the disease course of reactivated infections, then these trials suggest that the vaccine diminishes the incidence of HPV lesions by blocking true reinfection. Vaccinating children before sexual debut clearly reduces infection rates (4), but if vaccination prevents autoinoculation, then vaccinating previously infected individuals may also reduce HPV prevalence.

The type-specific effects of demographic and behavioral risk factors that we observed suggest that modest differences exist in the host subpopulations supporting each HPV type. 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 (3, 29). Smoking can suppress mucosal and cellular immunity (58), but its effects may be confounded with other high-risk behaviors that were not documented in the HIM Study. Although shared risk factors account for most of the initial infection risk for each type, there are important distinctions too. For instance, circumcision has a type-specific effect on the risk of HPV infection. Randomized, controlled trials of male circumcision have shown that circumcision protects against any HPV infection (59), whereas observational studies have documented either no effect or increased risk (3, 29). These conflicting results may have arisen from differences between types, potentially reflecting adaptation to different epithelial environments in the male genital tract.

Our modeling approach has several limitations: notably, that we cannot distinguish between reactivation and autoinoculation and that unmeasured covariates could have affected our results. While we identified celibate individuals based on detailed self-reporting on anal, vaginal, and oral sex, celibacy did not account for all forms of nonpenetrative sexual contact, such as kissing and masturbation. The construction of covariates from risk factor data, a necessary feature of any quantitative study of individual-level risk factors, may have led to bias (SI Appendix), and we cannot completely rule out the possibility that unobserved properties of the sexual contact network affected our results. In addition, our model of homologous immunity assumes that protection arises immediately after infection, but protection may be lagged (32). The poor fit of the homologous immunity model nevertheless suggests that any such effects are weak, but future investigations should allow for more complex forms of homologous immunity. We lack information on genetic differences in host susceptibility and on HPV exposure or changes in risk factors that occurred before the study. For example, we would have little power to detect very small effects in the few individuals who developed sterilizing immunity before entering the study. Finally, the data track genital HPV infection, 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. Our results imply that the HPV vaccine has the potential to lower HPV prevalence far more than previously expected.

### Materials and Methods

#### Data

The HIM Study is a multinational cohort study of HPV infection in unvaccinated men ages 18–70 y old. The study enrolled 4,123 participants between 2005 and 2009 and tracked men longitudinally over 5 y. Men were recruited from three cities: Tampa, Florida; Cuernavaca, Mexico; and Sao Paulo, Brazil. Detailed study methods are described elsewhere (36). The human subjects committees of the University of South Florida, Tampa, Florida; Centro de Referência e Treinamento em Doenças Sexualmente Transmissíveis e AIDS, Sao Paulo, Brazil; and Instituto Nacional de Saúde Pública de México, Cuernavaca, Mexico approved all HIM Study procedures. All participants provided written informed consent. The data for each individual consist of binary time series of infection status with each HPV type for up to 10 clinic visits (median = 10 visits). More information is in SI Appendix.

#### Model of HPV Dynamics

Our models are two-state discrete time partially observed Markov processes, in which an individual is either infected (one) or uninfected (zero) at any time with an HPV type. Infection of individual i with HPV type j occurs at rate λij, such that the probability of infection over Δt is

$$p(1+\Delta t|0) = 1 - e^{-\lambda ij \Delta t}. \quad [5]$$

Because we cannot directly observe infection events, the measurement model relates each observation Yijt to the latent state Xijt:

$$Y_{ijt} \sim \text{Bernoulli}(X_{ijt}(1 - P_0) + (1 - X_{ijt})P_H), \quad [6]$$

where P0 and PH are the rates of false positives and false negatives, respectively. The duration of infection was drawn from a gamma distribution, where the shape (k) and scale (ξ) were fixed to match the empirical distribution of infection durations in the data for type j (SI Appendix, Fig. S8 and Table S7).

#### Statistical Inference and Parameter Estimation

We used an iterative filtering (maximum likelihood via iterated filtering) (40) implemented in the R
package panelPomp (version 0.3.1), an extension of the pomp package (60).


Acknowledgments. Ken Alexander was instrumental in the project initiation. We thank Dan Zinder for helpful discussions about the model construction. The University of Chicago Research Computing Center provided valuable resources. Funding was provided by NIH Grants F30AI124636 (to S.L.R.) and T32GM007281.

6 of 6 | www.pnas.org/cgi/doi/10.1073/pnas.1714712114

Ranjeeva et al.