Integrating measures of viral prevalence and seroprevalence: a mechanistic modeling approach to explaining cohort patterns of HPV in women in the U.S.

Supplementary material

Population characteristics

Here, we demonstrate that the subpopulation of 6,442 female participants considered in this analysis do not differ substantially from the more general population of 8,148 female participants who may or may not have been vaccinated for HPV and may or may not have had conclusive genital HPV or HPV antibody assays. We consider a sample of demographic and reproductive and sexual health metrics in Table S1.

Table S1: Comparison of demographic and reproductive and sexual health metrics between all female participants ages 18–59 in 2003–10 and the subpopulation considered in this study.

Variable	Full population	Subpopulation
Number of participants, N	8148	6442
Age	38.5 (38.1–38.9)	39.2 (38.9–39.7)
Race		
Non-Hispanic White	66.4% (62.7–70.0%)	67.7% (63.9–71.4%)
Non-Hispanic Black	13.2% (11.2–15.2%)	12.5% (10.4–14.6%)
Mexican America	8.7% (7.0–10.5%)	9.0% (7.0–10.9%)
Other Hispanic	5.0% (3.9–6.1%)	5.0 (3.8–6.1%)
Other/Multiracial	6.7% (5.6–7.9%)	5.9 % (5.0–6.9%)
Ever taken birth control pills?	78.4% (76.9–80.0%)	80.0% (78.4%–81.5%)
Ever taken female hormones?	15.6% (14.4–16.9%)	16.5% (15.2–17.9%)
Ever had vaginal, oral, or anal sex?	96.0% (95.3–96.6%)	96.7% (96.1–97.3%)
Age at sexual debut?	18.6 (17.8–19.5)	18.8 (17.8–19.8)
Always used condoms in past year?[†]	29.9% (27.9–32.0%)	32.8% (30.6–35.5)
Number of lifetime male sex partners	8.3 (7.8–8.8)	8.4 (7.8–9.0)
Number of lifetime female sex partners	0.6 (0.2–1.0)	0.6 (0.1–1.1)

[†]: Excludes 2003–04

Identifiability of disease model

Here, we demonstrate that the disease model with data corresponding to I_{neg}, I_{pos}, and

$$\begin{aligned} W_{\text{neg}} &:= S_{\text{neg}} + L_{\text{neg}}, \\ W_{\text{pos}} &:= S_{\text{pos}} + L_{\text{pos}}, \end{aligned} \tag{S1}$$

is identifiable using differential algebra methods. For a more in-depth description of the differential algebra approach to structural identifiability and its theory and methods, we refer the reader to [S1–S4]. The system of six differential equations is

$$\begin{aligned} \dot{S}_{\text{neg}} &= \gamma I_{\text{neg}} + \omega S_{\text{pos}} - \lambda(c,a) \cdot S_{\text{neg}}, \\ \dot{I}_{\text{neg}} &= \lambda(c,a) \cdot S_{\text{neg}} + \omega I_{\text{pos}} + \mu L_{\text{neg}} - (\gamma + \sigma + \nu) I_{\text{neg}}, \\ \dot{L}_{\text{neg}} &= \nu I_{\text{neg}} + \omega L_{\text{pos}} - \mu L_{\text{neg}}, \\ \dot{S}_{\text{pos}} &= \gamma I_{\text{pos}} - \omega S_{\text{pos}} - \rho\lambda(c,a) \cdot S_{\text{pos}}, \\ \dot{I}_{\text{pos}} &= \rho\lambda(c,a) \cdot S_{\text{pos}} + \sigma I_{\text{neg}} + \mu L_{\text{pos}} - (\gamma + \omega + \nu) I_{\text{pos}}, \\ \dot{L}_{\text{pos}} &= \nu I_{\text{pos}} - (\mu + \omega) L_{\text{pos}}. \end{aligned} \tag{S2}$$

Through a series of substitutions, we can convert this system of equations into a series of four equations that uses only the variables corresponding to the observed data.

$$\begin{aligned} 0 &= \ddot{I}_{\text{neg}} - \mu\dot{W}_{\text{neg}} - \omega\left(\frac{\lambda+\mu}{\lambda\rho-\mu} + 1\right)\dot{I}_{\text{pos}} + (\gamma + \lambda + \nu + \sigma)\dot{I}_{\text{neg}} \\ &\quad + \mu\omega\left(\frac{\lambda-\mu}{\lambda\rho-\mu}\right) W_{\text{pos}} + \left(\frac{\omega(\lambda(\gamma+\mu-\lambda\rho-\nu-\omega)+\mu(\gamma+\nu+\omega))}{\lambda\rho-\mu}\right) I_{\text{pos}} \\ &\quad + \left(\lambda(\nu+\sigma) - \mu\gamma + \sigma\omega\left(\frac{\lambda-\mu}{\lambda\rho-\mu}\right)\right) I_{\text{neg}}, \\ 0 &= \dot{W}_{\text{neg}} + \dot{I}_{\text{neg}} - \omega(W_{\text{pos}} + I_{\text{pos}}) + \sigma I_{\text{neg}}, \\ 0 &= \dot{W}_{\text{pos}} + \dot{I}_{\text{pos}} + \omega(W_{\text{pos}} + I_{\text{pos}}) - \sigma I_{\text{neg}}, \\ 0 &= \ddot{I}_{\text{pos}} - \mu\dot{W}_{\text{pos}} + (2\omega + \lambda\rho + \gamma + \nu)\dot{I}_{\text{pos}} - \sigma\dot{I}_{\text{neg}} - \mu(\lambda\rho + \omega) W_{\text{pos}} \\ &\quad + (\gamma\mu + \lambda\rho\nu + \omega(\gamma + \nu + \lambda\rho + \omega) I_{\text{pos}} - \sigma(\lambda\rho + \omega) I_{\text{neg}}. \end{aligned} \tag{S3}$$

These equations are input–out equations for this system, i.e., they are monic, differential polynomial equations of the variables corresponding to the observed data. From these equations we

extract the all coefficients that are unique up to a real number:

$$\begin{aligned} C =&\{1, \mu, \lambda\mu, \sigma, \gamma + \lambda + \nu + \sigma, \omega, \mu\omega \left(\frac{\lambda - \mu}{\lambda\rho - \mu}\right), \omega \left(\frac{\lambda + \mu}{\lambda\rho - \mu} + 1\right), \mu(\lambda\rho + \omega), \sigma(\lambda\rho + \omega), \\ & \gamma + \nu + \lambda\rho + 2\omega, \lambda(\nu + \sigma) - \mu\gamma + \sigma\omega \left(\frac{\lambda - \mu}{\lambda\rho - \mu}\right), \\ & \frac{\omega(\lambda(\gamma + \mu - \lambda\rho - \nu - \omega) + \mu(\gamma + \nu + \omega))}{\lambda\rho - \mu}, (\gamma\mu + \lambda\rho\nu + \omega(\gamma + \nu + \lambda\rho + \omega)\}. \end{aligned} \tag{S4}$$

This set is the set of identifiable parameter combinations. The map $\{\mu, \lambda, \sigma, \gamma, \nu, \omega, \rho\} \to C$ is injective. Hence, all parameters are structurally identifiable. A mathematica file with all computations is available upon request.

Single-outcome models

The disease model fit to seroprevalence data uses the same model dynamics and all parameters are structurally identifiably (mathematica file available upon request). The genital-only model collapses to the following three-state model when $\rho = 1$. It is structurally identifiable.

$$\begin{aligned} \dot{S} &= \gamma I - \lambda(c, a) \cdot S, \\ \dot{I} &= \lambda(c, a) \cdot S + \mu L - (\gamma + \nu) I, \\ \dot{L} &= \nu I - \mu L, \end{aligned} \tag{S5}$$

The best-fit parameters for each model are compared in Table S2.

Table S2: Maximum likelihood parameter estimates and 95% confidence intervals for the joint genital–sero model, the genital-only model, and the sero-only model.

		Joint model		Genital-only model		Sero-only model	
Parameter	Definitions	Value	95% CI	Value	95% CI	Value	95% CI
$\lambda_0(1980)$	Force of infection coefficient	0.51	(0.43, 0.59)	0.46	(0.32, 0.61)	0.79	(0.47, 1.30)
γ	HPV cervicogenital clearance rate	0.41	(0.29, 0.52)	0.30	(0.15,0.46)	0.36	(0.11, 1.06)
σ	Seroconversion rate	0.74	(0.62, 0.86)	—	—	0.93	(0.45, 1.81)
ω	Rate of waning immunity	0.048	(0.035, 0.061)	—	—	0.07	(0.04, 0.13)
ν	Rate of entering latency	1.06	(0.75, 1.36)	0.31	(0.16, 0.47)	9.28	(4.90, 9.94)
μ	Rate of reactivation from latency	0.53	(0.28, 0.77)	0.21	(0.03, 0.38)	1.53	(0.71. 3.00)

References

[S1] Audoly, S., Bellu, G., D'Angiò, L., Saccomani, M. P., and Cobelli, C. (2001). Global identifiability of nonlinear models of biological systems. IEEE Transactions on Bio-medical Engineering, 48(1):55–65.

[S2] Meshkat, N., Anderson, C., and DiStefano, J. J. (2011). Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input-output equations. Mathematical Biosciences, 233(1):19–31.

[S3] Meshkat, N., Anderson, C., and DiStefano, J. J. (2012). Alternative to Ritt's pseudodivision for finding the input-output equations of multi-output models. Mathematical Biosciences, 239(1):117–123.

[S4] Saccomani, M., Audoly, S., Bellu, G., and D'Angio, L. (2001). A new differential algebra algorithm to test identifiability of nonlinear systems with given initial conditions. Proceedings of the 40th IEEE Conference on Decision and Control, 4:3108–3113.