Appendix

1.1 The unvaccinated model

The population of uninfected basal epithelial cells that HPV targets are represented by the variable X, and they are born at a rate $\lambda(t)$ and die naturally at rate μ. The population of free virions, V, come into contact with uninfected cells, X, and infect them at a rate ψ making infected cells, Y_1. Infection of new uninfected cells is limited by the fact that most cells are hidden under the epithelium and so abrasions are needed in order for HPV virions to reach them. For this reason we have slowed down the interaction between V and X by making their relationship grow hyperbolically (using a type-II functional response). Thus we assign the constant ϕ to be the density of uninfected cells at which the rate of growth of the Y_1 population is half-maximal.

These infected cells become self-replicating cells, Y_2, depending on the rate of oncogene expression, ε. Infected cells, Y_1 and Y_2, are killed by the CTL response, Z. The full model which includes all the assumptions mentioned in the methods is,

$$\begin{aligned}
\frac{dX}{dt} &= \lambda(t) - \mu X - \psi V\left(\frac{X}{\phi + X}\right) \\
\frac{dY_1}{dt} &= \psi V\left(\frac{X}{\phi + X}\right) - \varepsilon Y_1 - \mu Y_1 - aY_1Z \\
\frac{dY_2}{dt} &= \varepsilon Y_1 + r\varepsilon Y_2 - \mu Y_2 - aY_2Z \\
\frac{dV}{dt} &= \mu(k_1Y_1 + k_2Y_2) - \delta V \\
\frac{dZ}{dt} &= \omega Y_2 Z
\end{aligned} \qquad \text{(A.1)}$$

To reduce this model, we assumed that birth rate of the uninfected cells, $\lambda(t)$, maintains the total population size of epithelial cells at a constant population size of N and $\frac{d(X+Y_1)}{dt} = 0$, thus X can be replaced by $X = N - Y_1$. Thus, the Y_1.equation becomes

$$\frac{dY_1}{dt} = \psi V\left(\frac{N - Y_1}{\phi + (N - Y_1)}\right) - \varepsilon Y_1 - \mu Y_1 - aY_1Z \qquad \text{(A.2)}$$

as seen in model 1 in the methods.

1.2 Simplified model

We considered a simpler version of this model that only contained one class of infected cells,

$$\begin{aligned}
\frac{dX}{dt} &= \lambda(t) - \mu X - \psi V\left(\frac{X}{\phi + X}\right) \\
\frac{dY}{dt} &= \psi V\left(\frac{X}{\phi + X}\right) + r\varepsilon Y - \mu Y - aYZ \\
\frac{dV}{dt} &= \mu kY - \delta V \\
\frac{dZ}{dt} &= \omega YZ
\end{aligned} \tag{A.3}$$

Here, the *Y* equation grows either by the infection of uninfected cells (first term) or from its own self-division (second term). The results of this model were very similar to the one in the text, with two main exceptions. The unvaccinated immunity does not select for a low oncogene expression (Fig. A1 a) but when connected to the partnership model, the transmission constraints select for a low oncogene expression (Fig. A1 b). This shows how within- and between-host selection pressures can be at odds, and, in this case, the between-host selection pressure determines the optimal strategy. The other main difference is that the vaccinated short partnership behavior group requires significantly higher oncogene expression (than the super-spreaders and causal groups) to allow persistent circulation in this sexual behavior group (Fig. A1 d).

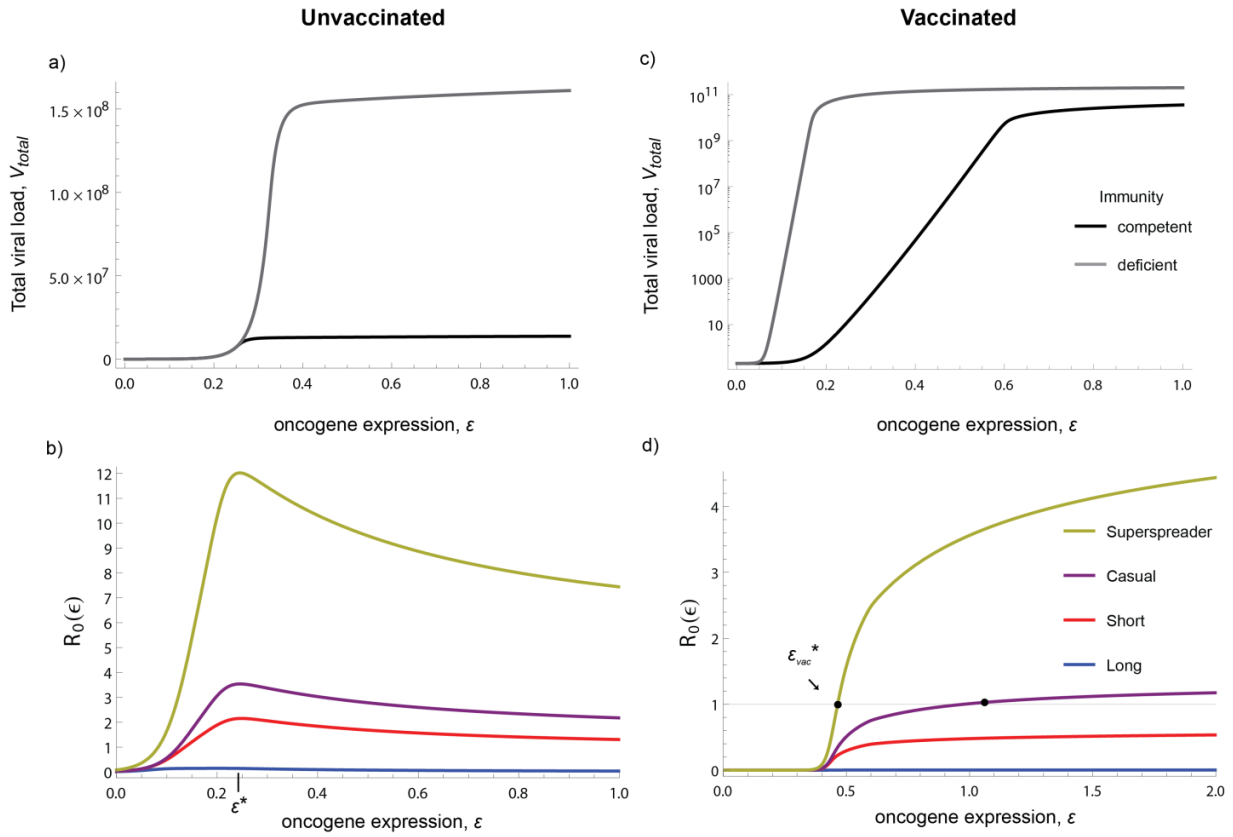

Figure A1. Unvaccinated and vaccinated within-host V_{total} (a and c respectively) and unvaccinated and vaccinated between-host selection for optimal oncogene expression (b and d respectively).

1.3 Sensitivity to parameter values

Attack rates

As an initial simplifying assumption, we considered the CTL attack rates against both Y_1 and Y_2 infected cell populations to be equal in strength (where Z removes either Y_i at a rate, a). However, to study the situation where CD8 T-cells attack the infected cell populations differentially, we considered slight alterations of models 1 and 2 such that a in equation dY_1/dt became a_1 and in equation dY_2/dt the attack rate specific to Y_2 cells become a_2. A biological reason for the natural immune response to exhibit differential attack rates would be that the increased oncogene expression in Y_2 were differentially targeted (otherwise, the two infected cell groups behave similarly). Indeed, the cell-mediated immune response needs to target E6 epitopes for effective clearance [1,2]. In this case, a_2 should be larger than a_1 because Y_2 cells maintain a higher oncogene expression. We considered this scenario, and found that even if a_2 was increased by 3 orders of magnitude (in relation to a_1) little changed. For instance, in the time-series the infected cells and viral load peak lower, a smaller population of CD8 T-cells are

needed to clear the infection and that the timing and the shapes of the curves remained the same (Fig. A2). Likewise, the ε^* values found by V_{total} and R_0 do not change compared to the scenario where the attack rates are the same (Fig. A3). The same can be said for the case where $a_1 > a_2$ (not shown), though a biological reason for this scenario is not apparent. Since the vaccine-induced immunity targets the L1 late protein, the two infected cell groups should be targeted at the same intensity by effector cells (as we considered in the main text). Nonetheless, we considered differential attack rates in vaccinated hosts for completeness. In the time-series, when $a_1 > a_2$ the Y_1 peak for the higher oncogene expression is lowered, while the rest of the curves stay almost the same (not shown). When $a_2 > a_1$, then all Y_1 curves decay and the growth of the Y_2 nearly instantaneous for higher oncogene expression values, thus the "increase rapid cell division before clearance" effect is more pronounced (Fig. A4). In both vaccine cases, less effector cells (lower Z) are needed to clear the infection and the V_{total} and R_0 give the same ε^* as when the attack rates are equal.

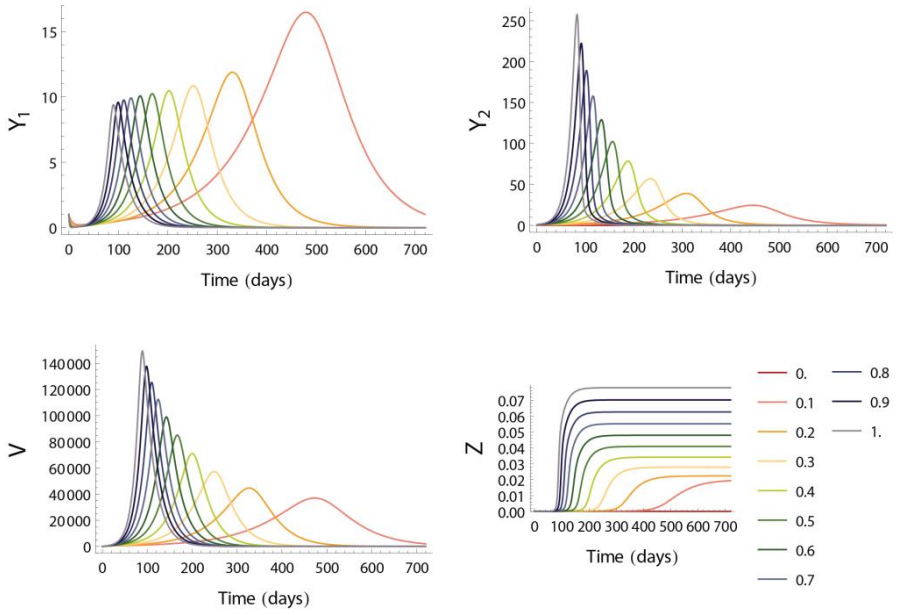

Figure A2. Unvaccinated time-series with $a_2 >> a_1$ ($a_2 = 10$ and $a_1 = 0.01$), for various ε values (from 0 to 1).

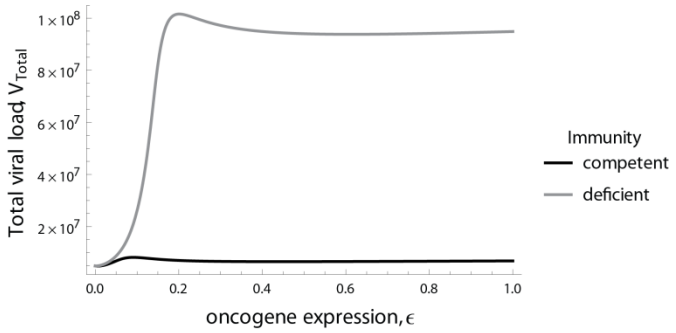

Figure A3. Unvaccinated V_{total} with $a_2 >> a_1$ (a_2 = 10 and a_1 = 0.01).

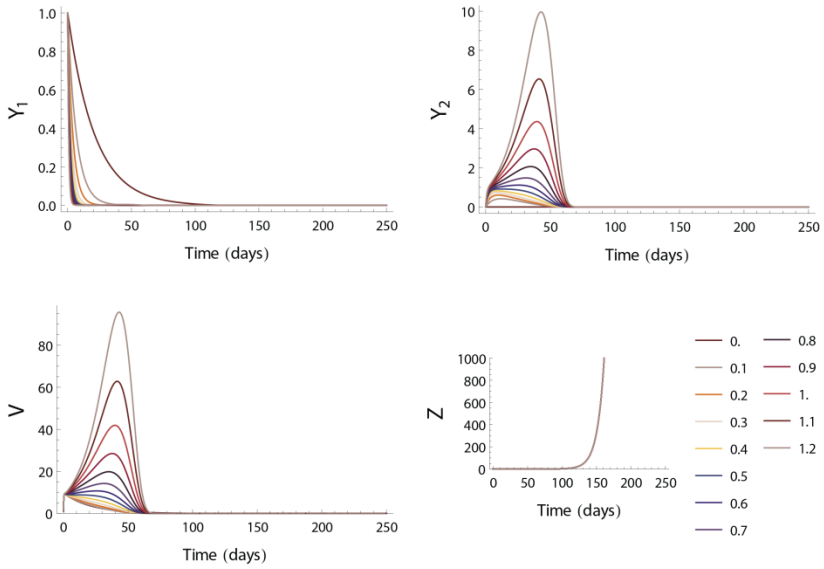

Figure A4. Vaccinated time-series with $a_2 >> a_1$ (a_2 = 10 and a_1 = 0.01), for various ε values (from 0 to 1.2).

Sexual behaviour parameters

It should be noted that partnership length and turnover can vary throughout a host's life. Therefore, HPV prevalence in different age groups should be linked to higher proportions of short or casual individuals in these age groups which permit more transmission of HPV. The average partner turnover at different age demographic groups (e.g. in 20s, 30s, or 40s) is cultural (and gender-specific), which might help explain variations in HPV prevalence in the same age groups across the world [3]. Indeed, we find that the parameter that most affects the host's R_0 is

partnership acquisition, ρ, which demonstrates that increasing the number of new partners (even when old partnerships have not broken up) increases the transmission of the virus (e.g. fig. A5).

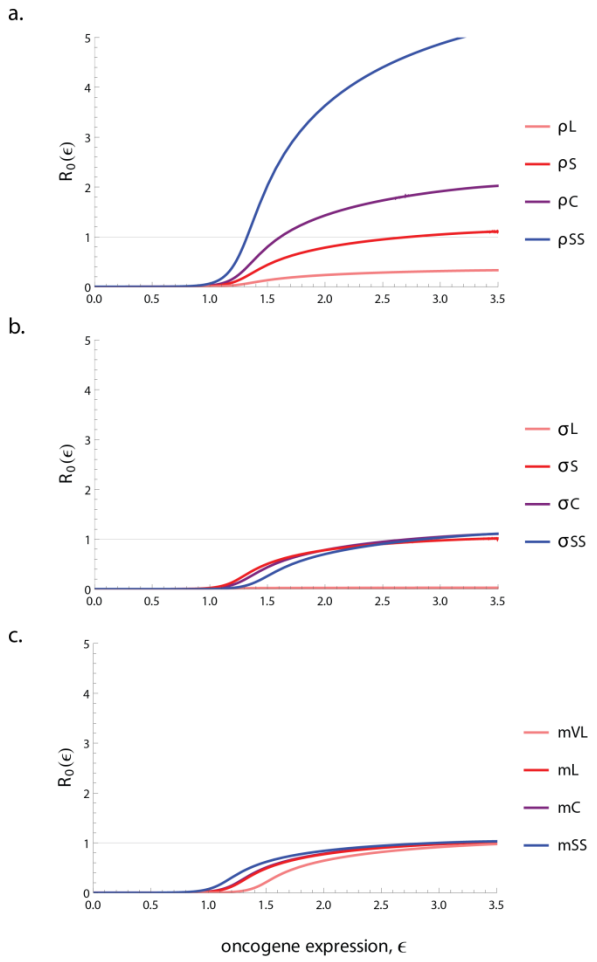

Figure A5. Vaccinated hosts: oncogene expression needed for transmission, ε_{vac}*, is the point where the curve crosses $R_0 = 1$, and is most affected by acquisition of new partners rate, ρ. **a)** parameters $\sigma = 0.05$ and $m = 0.43$ are held constant, and ρ is varied ($\rho L = 0.0027$, $\rho S = 0.0096$, $\rho C = 0.019$, $\rho SS = 0.068$). **b)** paramters $\rho = 0.0096$ and $m = 0.43$ are held constant and σ is varied ($\sigma L = 0.0004$, $\sigma S = 0.05$, $\sigma C = 0.1$, $\sigma SS = 0.44$). **c)** parameters $\rho = 0.0096$ and $\sigma = 0.05$ are constant and m is varied ($mVL = 0.033$, $mL = 0.356$, $mC = 0.43$, $mSS = 1.45$).

References

1. Scott M, Nakagawa M, Moscicki A (2001) Cell-Mediated Immune Response to Human Papillomavirus Infection. Clin Vaccine Immunol 8: 209. doi:10.1128/CDLI.8.2.209.

2. Stanely MA (2006) Immunobiology of Papillomaviruses. In: Campo MS, editor. Papillomavirus Research: from natural history to vaccines and beyond. Horizon Scientific Press. p. 311.

3. Clifford GM, Gallus S, Herrero R, Muñoz N, Snijders PJF, et al. (2005) Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis. Lancet 366: 991–998. doi:10.1016/S0140-6736(05)67069-9.