**SUPPLEMENTARY MATERIAL**

**Model flow chart**



**Model equations**



with *i=1,2* for the sexual activity class (*1=low, 2=high)*, *j=1,2* for the smoking status (*1=non-smoker, 2=smoker), ni=* rate of new sexual partner acquisition, *Xi,j=* susceptible, *Yi,j=* infected, *Zi,j=* immune, *Ni,j=* total population (*Xi,j+Yi,j+Zi,j)*, =clearance rate, =mortality rate, =probability of transmission per partnership, =probability of developing natural immunity after clearance, *pi,k(j,l)*=proportion of partnerships of individuals in class *(i,k)* that are with individuals in class *(j,l)*.

**Mixing equations**

The quantities *pi,k(j,l)* are the proportion of partnerships of individuals in the class *(i,k)* that are formed with individuals in the class *(j,l)* and are given by the mixing equations taken from [[1](#_ENREF_1)]:

 (1)

Where *a,b,c* are non-negative numbers such that *a+b+c*≤*1* and is equal 1 if *i=j* and 0 otherwise. Other notation is as in Web Appendix 1.

These mixing equations are algebraically always balanced for any non-negative *a,b,c* such that *a+b+c*≤*1* because they are a convex sum of solutions (each term of the sum can be seen to be a solution).

Formula (1) can be understood by thinking of partnership formation as an individual having three different types of partner selection: each term of the sum represents a different type of partner selection and the parameters *a, b, c, 1-a-b-c* are the probabilities of each type of selection. The first term is the event of selecting assortatively a partner from the same smoking and sexual activity class. The second term is selecting from the same sexual activity class, but randomly as for smoking status. Thus, if the second choice is made, it is still possible to be choosing from the same sexual activity and smoking class. The third term is random selection for sexual activity, and assortatively for smoking. The last is selecting completely at random. Thus, the *assortativity bias* comes from the third term, where the probability of selecting a highly sexually active individual is greater for smokers.

Sticking with the interpretation of formula (1) as probabilities of event, we can define the four probabilities as obtained by two probabilities: the probability of selecting assortatively for smoking status, *ε1*, and for sexual activity, *ε2*. The alternative to selecting assortatively is selecting randomly. Thus, the probability *a* becomes *ε1ε2.* We can further introduce a correlation between the two events using a parameter, *c* (*c* is not the correlation however). The new formula would then be:



**Adjustment for sexual activity**

Keeping the same notation as in Web Appendix 1, the formula for adjustment for sexual activity was:



**REFERENCES**

(1) **Walker R, et al.** A revision of sexual mixing matrices in models of sexually transmitted infection. *Stat Med* 2012; **31**(27): 3419-3432.

(2) **Van de Velde N, et al.** Population-level impact of the bivalent, quadrivalent, and nonavalent human papillomavirus vaccines: a model-based analysis. *J Natl Cancer Inst* 2012; **104**(22): 1712-1723.

(3) **Drolet M, et al.** The psychosocial impact of an abnormal cervical smear result. *Psychooncology* 2012; **21**(10): 1071-1081.