ASSORTATIVE MIXING AS A SOURCE OF BIAS IN EPIDEMIOLOGICAL STUDIES OF SEXUALLY TRANSMITTED INFECTIONS: THE CASE OF SMOKING AND HUMAN PAPILLOMAVIRUS

Lemieux-Mellouki P, Drolet M, Brisson J, Franco EL, Boily M-C, Baussano I, Brisson M

ABSTRACT

For studies examining risk factors of sexually transmitted infections (STIs), confounding can stem from the characteristics of the partners of study subjects, and persist after adjustment for the subjects' individual-level characteristics. Two conditions can produce confounding by the subjects' partners: C1) partner choice is assortative by the risk factor examined and, C2) sexual activity is associated with the risk factor. We term this bias the *assortativity bias*. The objective of this paper is to illustrate the potential impact of the *assortativity bias* in studies examining STI risk factors, using smoking and Human papillomavirus (HPV) as an example. We developed an HPV transmission-dynamic mathematical model in which we nested a cross-sectional study assessing the smoking-HPV association. In our base-case, we assumed 1) no effect of smoking on HPV, and 2) conditions C1-2 hold for smoking (based on empirical data). The *assortativity bias* caused an overestimation of the odds ratio (OR) in the simulated study after perfect adjustment for the subjects' individual-level characteristics (adjusted OR=1.51 instead of 1.00). The bias was amplified by a lower basic reproductive number (R_0), greater mixing assortativity and greater sexual activity among smokers. To mitigate the *assortativity bias* studies should adjust for the characteristics of study partners.

Infectious diseases are transmissible, which implies that the contact network of an individual is a crucial determinant of his risk of infection (1). Thus, an individual's risk of infection not only depends on individual-level risk factors, such as his age and gender, but on network-level risk factors. A classic example of this particularity of infectious diseases is herd immunity: vaccinating a portion of the population reduces the chance of non-vaccinated individuals being exposed to the infectious agent (2). Hence, an individual's risk of infection following the introduction of vaccination depends on his/her vaccine status (individual-level risk factor), and the overall population-level vaccination coverage (network-level risk factor).

In studies that examine the biological causes of sexually transmitted infections (STIs), control of confounding habitually follows the traditional non-transmissible disease approach of controlling for individual-level risk factors, with little attention to control for network-level risk factors such as the characteristics of the individual's partners. In doing so, authors usually acknowledge the possibility of misclassification of sexual activity, due to misreporting by study subjects, which can cause residual confounding of effect measures if sexual activity is a confounder (3, 4). However, what is not generally acknowledged is that, even if the sexual activity of the study subject was perfectly measured and controlled for, the sexual activity of the partners of study subjects can also cause confounding. We term one particular form of confounding bias the *"assortativity bias"*, because it stems from assortativity in sexual mixing (partner choice). That is, on average, people have partners with similar characteristics (e.g., age) and behavior (e.g., smoking status) as themselves (5-8).

In STI studies, *assortativity bias* can occur if two conditions are met: (C1) partner choice is assortative according to the risk factor of interest, and (C2) the association between this factor and infection is confounded by sexual activity. When these conditions are met, the risk factor of interest can be associated with the likelihood of having an infected partner, and a confounding bias is likely to remain even after perfectly controlling the effect measure for the individual-level confounders.

To illustrate the *assortativity bias*, we consider a simplified example where the effect of smoking on the risk of a STI is examined. In this example, *assortativity bias* can occur because the two conditions above are met: sexual mixing is assortative according to smoking status (C1) (6, 8), and the relation of smoking with the STI is confounded by sexual activity (smokers have more lifetime sexual partners than non-smokers and number of lifetime partners is a risk factor of STI (C2))(9, 10).

Figure 1 illustrates the essential components of *assortativity bias*, assuming, for simplicity, that smoking is the only factor for which partner selection is assortative and that smoking itself has no biological effect on STI acquisition/transmission or duration. In Figure 1, the study subject's smoking status is positively associated with the smoking status of his partner (assortative mixing by smoking status (C1)). A subject who smokes will be at greater risk of infection not only because of his own sexual activity, but also because his/her partner is likely to be a smoker and thus more likely to be more sexually active and infected. The smoking-STI association is confounded by sexual activity of the subjects (Figure 1 right panel, dotted arrows) and subjects' partners (Figure 1 left panel, dotted arrows). Therefore, even if sexual activity of the subjects is controlled for (individual-level), residual confounding bias remains possible due to sexual activity of partners (network-level). Even if we assume a direct causal

effect of smoking on STI risk, the *assortativity bias* would still occur resulting in an overestimation of the magnitude of the effect for smoking on STI risk. Of note, the *assortativity bias* has a dynamical aspect: having a partner that smokes not only induces greater risk through the partner's greater chance of being highly sexually active, but also through the partner's own previous partners who were also more likely to be smokers. Conditions 1-2 are met for many common risk factors of STIs, including age, race and socio-economic status (SES), and therefore *assortativity bias* is likely present in many prospective and cross-sectional epidemiological STI studies.

Mathematical modeling has been used to understand potential biases in epidemiological studies of STIs (11-14). With modeling, an artificial world is created where transmission and natural history of disease can be simulated based on observed data. Epidemiological studies can be nested within the model to examine the presence/magnitude of potential biases under different assumptions regarding behavior, transmission, and natural history.

In this paper, we examine the *assortativity bias* focusing on the smoking–STI association. Smoking is a suspected risk factor for STIs through increased transmission and/or duration of infection (15). As smoking is a modifiable behavior, there is a high interest in understanding its role on STI incidence/prevalence. Smoking has been independently associated with the prevalence of STIs such as human immunodeficiency virus (HIV), herpes simplex virus 2 (HSV-2), genital and oral infections by human papillomavirus (HPV) in many studies (4, 9, 16-18), and these associations have been shown to follow monotone dose-response relationships (9). However, the possibility of *assortativity bias* was not addressed or controlled for in these studies.

The specific objectives of this paper are to use mathematical modeling: 1) to illustrate and describe the *assortativity bias*, using the association between smoking and genital HPV infection, and 2) to examine the sensitivity of the *assortativity bias* to the biological and behavioral parameters, for generalization of results.

METHODS

Mathematical model

We developed a deterministic transmission dynamic model of HPV infection (see Web Appendix 1 for equations). The modeled population is heterosexual, open and stable. For the base-case scenario, we modeled HPV16 infection, which is the most prevalent type and causes the majority of HPV-related cancers. The simulated population is stratified for the two behavioral aspects from which the the *assortativity bias* stems: 1) smoking status (smoker/non-smoker), and 2) sexual activity (low/high). For simplicity, we did not stratify the model by age. On average, individuals spend 30 years in the modeled population, representing the years of higher sexual active life (e.g., 15-44 years of age). Sexual mixing depends on an individual's smoking status and sexual activity. For each of these behavioral factors, we allowed mixing to vary from random to completely assortative (i.e., smokers only form partnerships with smokers). Hence, the model corresponds to Figure 1, but adding an arrow between the sexual activity of subjects and sexual activity of their partners. We included assortativity according to sexual activity as it is a key feature of sexual networks (19, 20). Based on empirical evidence, we assume that the two conditions for the *assortativity bias* are met: C1) sexual mixing is assortative for smoking (6, 8), and C2) smokers are more sexually active than non-smokers (9, 10).

Importantly, in our base-case, we assumed that smoking is not a cause of infection, to test whether observed associations between smoking and HPV infection can be explained by the *assortativity bias*. More specifically, we assumed that smoking does not increase HPV transmission probabilities or duration of infection. We also tested the extent of the bias under the assumption of a true biological effect of smoking on HPV infection (see below under sensitivity analyses).

Parameterization

Model parameter values and references are presented in the Web Table 1. We used biological parameter values estimated in prior modeling work (21-23), and estimated the proportion of smokers in each sexual activity class from an epidemiological study (24). Although studies suggest that sexual mixing by level of sexual activity and smoking status is assortative (6, 8, 25, 26), no empirical estimates are available in the literature. In our base-case, we assumed assortativity parameters values for smoking status and for sexual activity to be 0.8 and 0.4, respectively (0.0=no assortativity, 1.0=complete assortativity), using parametric equations for mixing presented in the Web Appendix 2. We performed extensive sensitivity analysis on assortativity parameter values given their uncertainty.

Experimental design and outcome measure

To examine the association between smoking and HPV, we nested a prevalence study in the simulated population. Study subjects are a cross-section of the entire simulated population. HPV prevalence among study subjects was estimated at endemic equilibrium without HPV vaccination. We used odds ratios (ORs) of HPV infection (positivity) among smokers compared to non-smokers as the measure of association. The overall adjusted ORs were calculated as the weighted average of the stratum-specific ORs of the two sexual activity classes using, as

weights, the proportion of the population in each sexual activity class (see Web Appendix 3 for equations). In the simulated study, <u>adjusted or stratum-specific ORs different from 1.00 can only be due to *assortativity bias*, because there is no biologic effect of smoking on infection in our model</u> and sexual activity of subjects is perfectly adjusted for. The magnitude of the *assortativity bias* is the magnitude of the deviation of the adjusted and stratum-specific ORs from 1.00.

Sensitivity analyses

The aim of the sensitivity analyses was to 1) assess the impact of parameter uncertainty on the magnitude of the *assortativity bias*, and 2) illustrate how the *assortativity bias* could change for different STIs and population-level behaviors. We varied the key biological/behavioral parameters, one at a time, keeping the value of all other parameters fixed at their base-case values. We used the stratum-specific ORs to isolate the bias in each sexual activity class. Finally, we estimated the potential impact of the *assortativity bias* on adjusted OR assuming different magnitudes of a true biological effect of smoking on HPV infection.

RESULTS

Base-case

Table 1 shows the base-case model predictions of the ORs of HPV infection among smokers compared with non-smokers. We estimated crude and adjusted ORs of 1.64 and 1.51, respectively (Table 1). Given that, in our model, smoking has no causal effect on HPV and we can perfectly control for sexual activity of subjects (no residual confounding), *assortativity bias* is the only possible cause of adjusted OR>1.00.

The magnitude of *assortativity bias* is generally greater for those with lower sexual activity (Table 1, Figures 2-3). This is because highly sexually active subjects will likely have highly sexually active partners (assortativity by sexual activity), irrespective of smoking status.

Impact of behavioral factors

Association between sexual activity and smoking. The ORs of HPV infection (*assortativity bias*) increase as the strength of the association between smoking and sexual activity among study subjects increases (Figure 2A). This is because a stronger association causes greater confounding by sexual activity (increased impact of Condition 2 – Figure 1).

Assortativity by smoking status. Greater assortativity according to smoking status results in a steep increase in the ORs of HPV infection comparing smokers with non-smokers (Figure 2B). When smoking assortativity is stronger, the unbalance in sexual activity between smokers and non-smokers will be replicated between smokers' partners and non-smokers' partners to a greater extent (increased impact of Condition 1 – Figure 1).

Assortativity by sexual activity. The ORs of HPV infection decrease with greater assortativity between individuals of the same sexual activity class (Figure 2C). As assortativity by sexual activity increases, the sexual activity of the study subject becomes a better proxy of his partners' sexual activity. At the extreme, when mixing by sexual activity is completely assortative, subjects will have partners belonging to their own sexual activity class, irrespective of smoking status. Therefore, there will be no bias after stratifying for the sexual activity of study subjects.

Impact of biological factors

Transmission probability & duration of infectiousness. The model shows that the magnitude of the *assortativity bias* is highly sensitive to the transmission probability or duration of infection (Figure 3A-3B). The ORs of HPV infection among smokers compared with non-smokers, stratified by sexual activity of study subjects, decreases steeply with increased transmission probability or duration of infectiousness. In general, if the reproductive number is low (i.e. low transmission probability, short duration or low partner acquisition rate), the difference in sexual activity between smokers and non-smokers can lead to large differences in HPV prevalence between the two groups. For example, if smokers and non-smokers were two separate populations (i.e. complete assortativity), then there is a threshold where the probability of transmission is too low and the virus disappears from the non-smokers population, but is high enough for the virus to remain in the smokers' population, with greater sexual activity. In this extreme case, we would get an infinite OR.

Natural immunity. The probability of developing natural immunity has little impact on the magnitude of the *assortativity bias* (Figure 3C). This is because the relative reduction in HPV prevalence caused by lower natural immunity is identical for smokers and non-smokers.

Assortativity bias assuming a true biological effect of smoking

Figure 4 shows the OR when varying the effect of smoking on the duration of infection with and without assortativity by smoking status. The *assortativity bias* produces an overestimation of the OR when smokers have a longer duration of infection than non-smokers. This overestimation of OR rises steeply as the biological effect of smoking increases. The OR is also overestimated when smoking directly affects the transmission probability, or the probability of developing immunity after clearance (results not shown).

DISCUSSION

In this paper we present the *assortativity bias*, a frequently unrecognized confounding bias specific to studies examining risk factors of infectious diseases. To illustrate this bias, we considered the example of smoking as a possible biological cause of HPV infection. Using mathematical modeling, we showed that adjustment for the subjects' individual-level sexual activity is not enough to isolate the true biological effect of smoking when mixing is assortative by smoking status (C1) and smoking status is associated with sexual activity (C2). There is empirical evidence that these two conditions hold for smoking (6, 8-10), and many other risk factors of infectious disease such as age, race/ethnicity and SES. Hence, the *assortativity bias* is likely present in many epidemiological studies examining risk factors of STIs and other infectious diseases.

Our modeling results are consistent with the results from epidemiological studies examining the association between smoking and HPV or other STIs. In a recent large-scale study, the adjusted ORs of HPV infection in smokers compared to non-smokers was 1.4 (95%CI: 1.2, 1.7) (9), and most other studies have found ORs higher than 1.0 (18, 27-29). Our modeling analysis suggests that the *assortativity bias* could produce ORs of the magnitude seen in these empirical studies if assortativity by smoking status is high. The association between smoking and HPV infection is supported by traditional criteria of causality such as a dose-response. However, it should be noted that the *assortativity bias* can reproduce a dose-response relationship if (C1) mixing is assortative by smoking intensity and (C2) there is a dose-response relationship between sexual activity and smoking intensity. Significant associations between infection and smoking have also been observed in empirical-based studies for many other STIs (4, 16, 17). We showed that the size of the *assortativity bias* should be more important for STIs such as HIV, which have very small transmission probabilities. This is consistent with a

systematic review which showed that the association between HIV infection and smoking (4) was higher than for HPV (ORs between 1.2 and 3.5).

Our results should not be interpreted as evidence that smoking is not a cause of HPV infection. It is conceivable that smoking itself may have a direct biological influence on HPV risk by negatively affecting mucosal immunity and/or by consuming micronutrients that mediate resistance to or clearance of HPV infection (30). Our model served to illustrate the mechanisms behind the *assortativity bias*, and the role it can play in overestimating the ORs in studies examining smoking as a risk factor for STIs. When we assume, in our model, that smoking is a biological cause of HPV infection, the *assortativity bias* greatly increases the adjusted ORs beyond the true biological effect. In addition, it is important to note that the magnitude of the *assortativity bias* may vary between countries due to differences in behavior (differences in the magnitude of C1-2), and therefore associations between STIs and risk factors can vary substantially between studies.

The *assortativity bias* could affect many other risk factors than smoking, such as age and race. For example, young adults are generally the most at risk of STIs, even after adjustment for sexual activity (25, 31). It is suggested that this is due to a biological cause (cervical ectopy makes young women vulnerable to STIs) (32). Yet, sexual mixing is highly assortative with respect to age (5, 21, 33) (C1), and younger adults are more sexually active (33, 34) (C2), and hence an age-STI association can be partly due to the *assortativity bias*. Similarly, African Americans have higher STI prevalence after adjustment for sexual activity (7, 35). It has been shown that sexual mixing is assortative according to race/ethnicity (7, 20) (C1) and African Americans have younger age at sexual debut (36), which is a risk factor for many STIs (C2). For other risk factors, complete assortativity between individuals with the risk factor can hold

automatically and cause *assortativity bias* in prevalence studies. For example, being infected with HPV31 implies that a previous/current partner was infected with HPV31. Thus, there is *de facto* assortativity between individuals infected with HPV31 (C1). Because being infected with HPV31 is associated with sexual activity (C2), HPV31 can be measured in prevalence studies as a risk factor of other STIs, due to the *assortativity bias*.

The main strength of this study was the use of mathematical modeling to perfectly control a fictive population, allowing us to quantify the *assortativity bias*. It was thus possible to explore the theoretical basis for the bias and the relation between the bias and behavioral and biological parameters. However, the main limitation of our model is that many sources of heterogeneity (sexual activity, smoking intensity, mixing) were not included and we assumed independence between mixing by sexual activity and by smoking status. Particularly, we did not include in the model other factors that could cause assortativity by smoking status. For example, SES is a risk factor for smoking (37, 38), and sexual mixing is assortative by SES (38), which indirectly produces assortativity with respect to smoking. SES could thus partly explain the assortativity by smoking status. In this case, we would still observe a bias in the smoking-HPV association of the simulation, but this bias would be partly corrected by adjustment for the SES of study subjects. These model simplifications do not affect the robustness of our overall qualitative conclusions. However, in view of the model's simplicity, one should not use the precise numerical model estimates of the ORs as representative of reality.

To correct for the *assortativity bias* in studies examining smoking as a risk factor of HPV infection, one must control for systematic differences between smokers and non-smokers in exposure to infection. In prospective studies, the most direct method to avoid the bias is to assess the HPV status of new sexual partners of study subjects, and stratify the comparison of

HPV infection incidence between smoking and non-smoking subjects by exposure to infection (e.g., exposure to infected partners (39, 40)). Alternatively, the smoking status of new sexual partners of study subjects should be assessed to control for the higher chance of smokers having smoker partners. In our simulated study, adjusting for the smoking status of the subjects' partners and the subjects' sexual activity is sufficient to eliminate the *assortativity bias* (adjusted OR is 1.0). Given that many risk factors are investigated at once in empirical studies, it is necessary to have simultaneous adjustments for the key risk factors being investigated (e.g., age, race/ethnicity, SES) at the subject- and partner-level. However, real-world sexual networks are more complex than the one reproduced by our model. Hence, it is essential to include as many risk factors as possible to be used as proxies of HPV infection to control for exposure to infection of the subjects partners. Residual confounding will likely remain. Furthermore, risk factors of subjects' partners could also include the characteristics of their own partners (the partners of partners) which will likely not be known. Using proxies of characteristics of the whole sexual network of an individual, such as SES or the prevalence of smoking in the neighborhood, could complement the previous approach of adjusting for previous partners' individual-level characteristics. A possible statistical framework for this would be multi-level model (41). In a cross-sectional study, information on past partners will be needed, with a recall window being a function of the duration of infection. Clearly, more research is required to identify the key partner and network characteristics that should be included in studies, and, in the case of cross-sectional studies, the number of past partners needed to reduce the magnitude of residual confounding.

In conclusion, assortative sexual mixing by smoking status can cause bias in studies assessing the biological effect of smoking on HPV acquisition. This *assortativity bias* is likely to occur for other risk factors and infectious diseases. For a thorough adjustment of measures of

association, data on risk factors of sexual partners of study subjects is required to mitigate the impact of the bias.
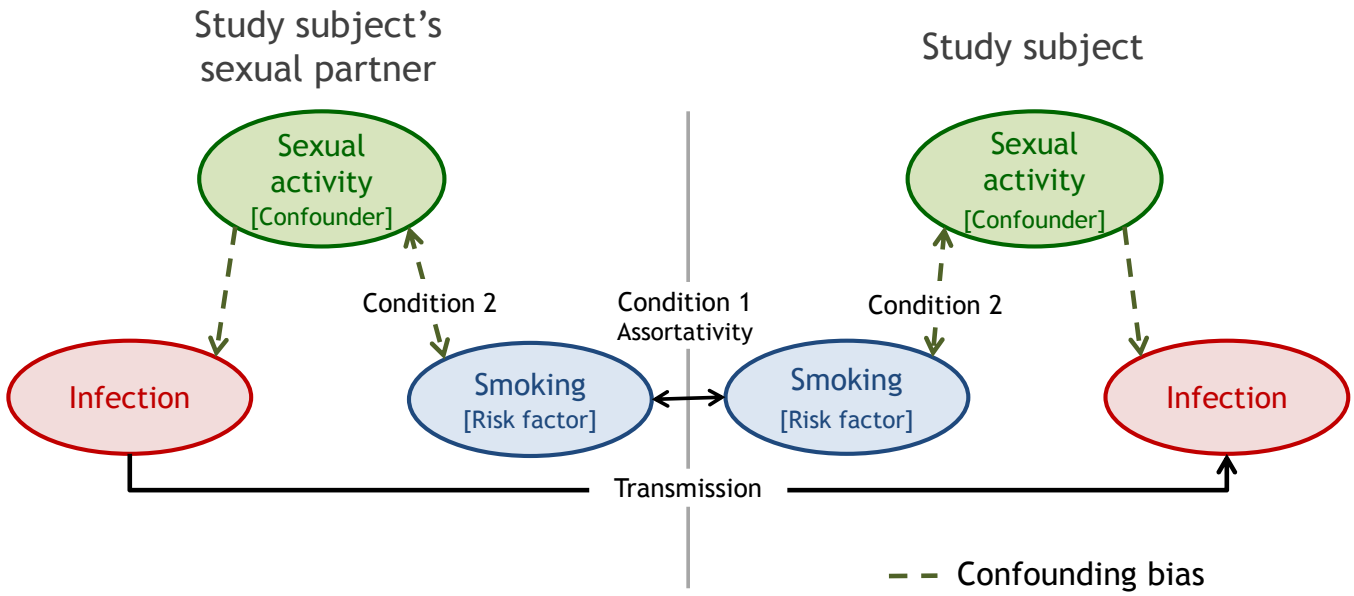
ACKNOWLEDGEMENTS

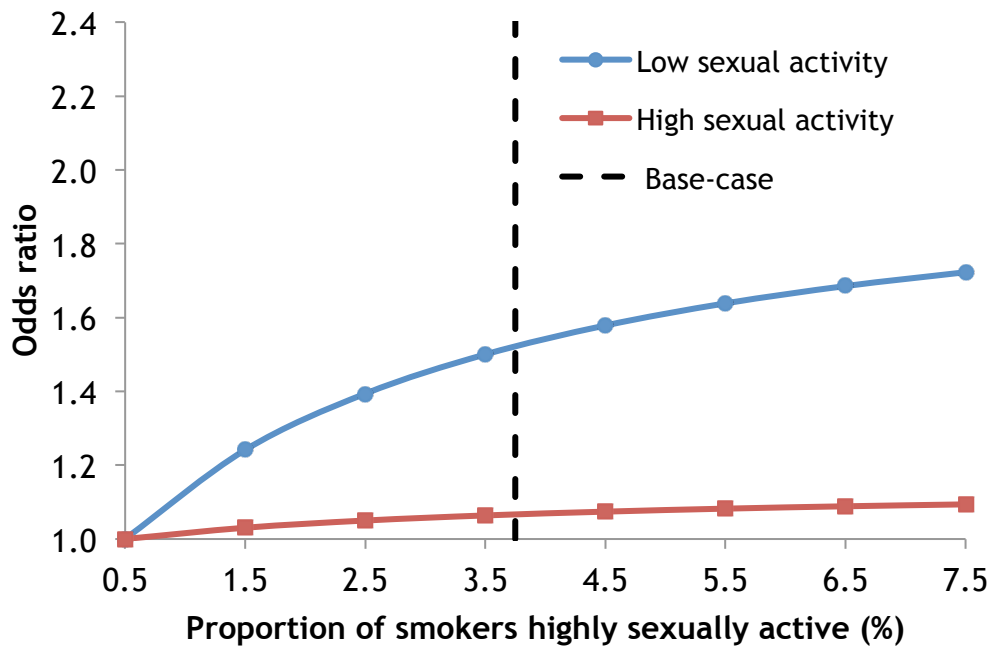**Figure 1**

Table 1. Odds Ratios of HPV Infection Between Smokers and Non-Smokers Among Modeled

Study Subjects

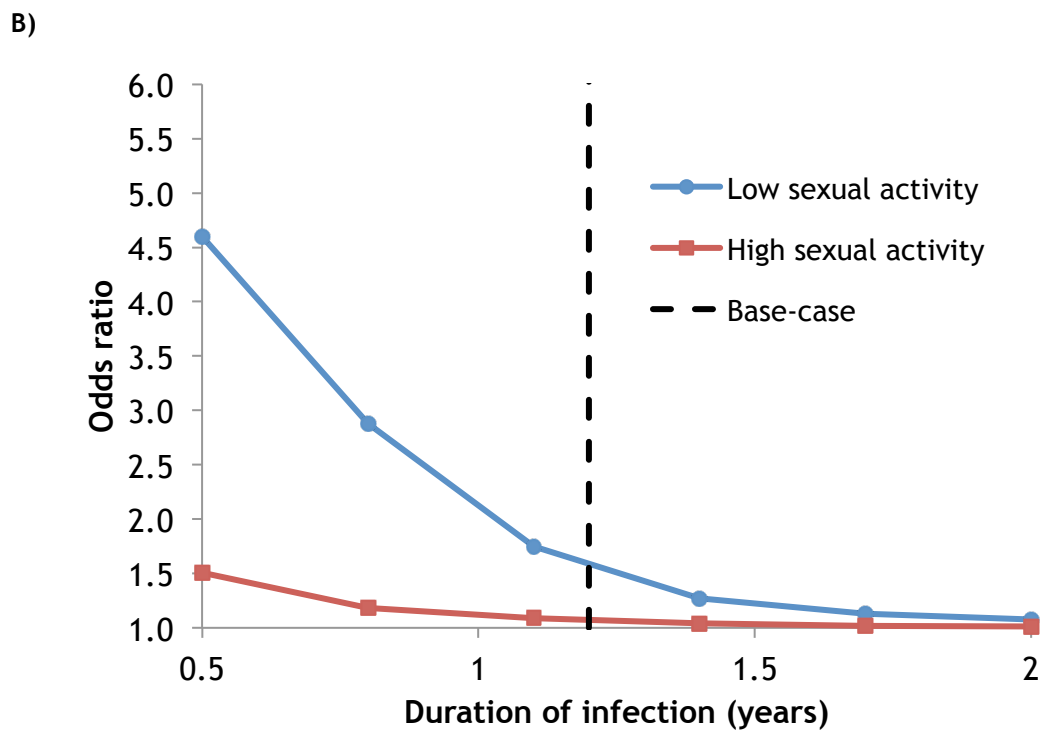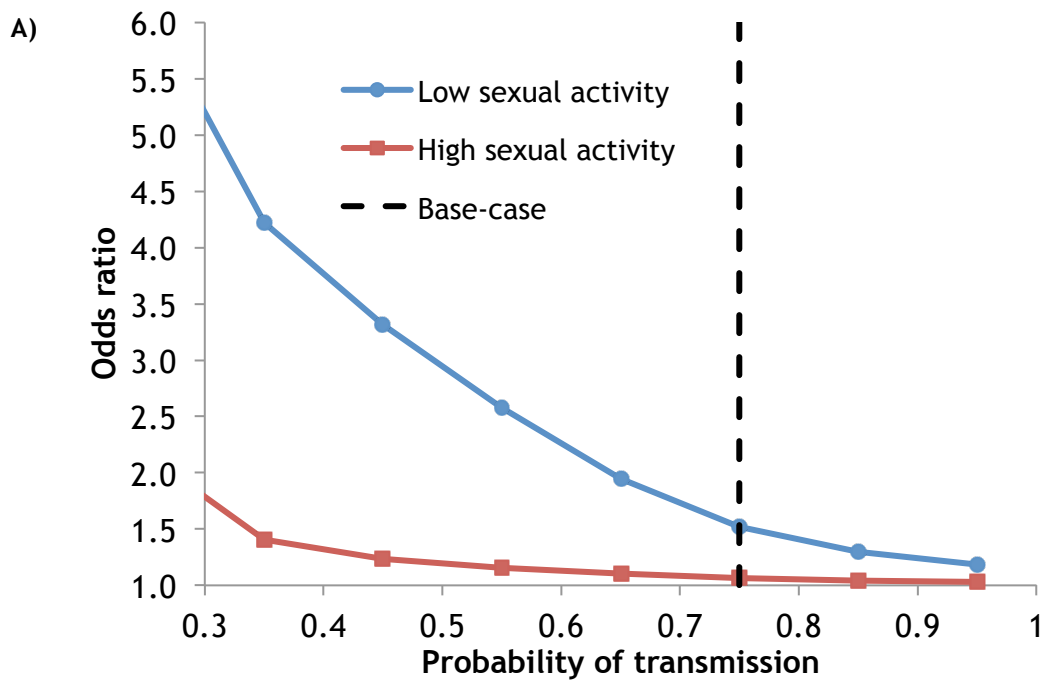|	Odds ratio
Crude	1.64
Stratified	
Low sexual activity class	1.52
High sexual activity class	1.07
Adjusted	1.51

Figure 2

A)

B)
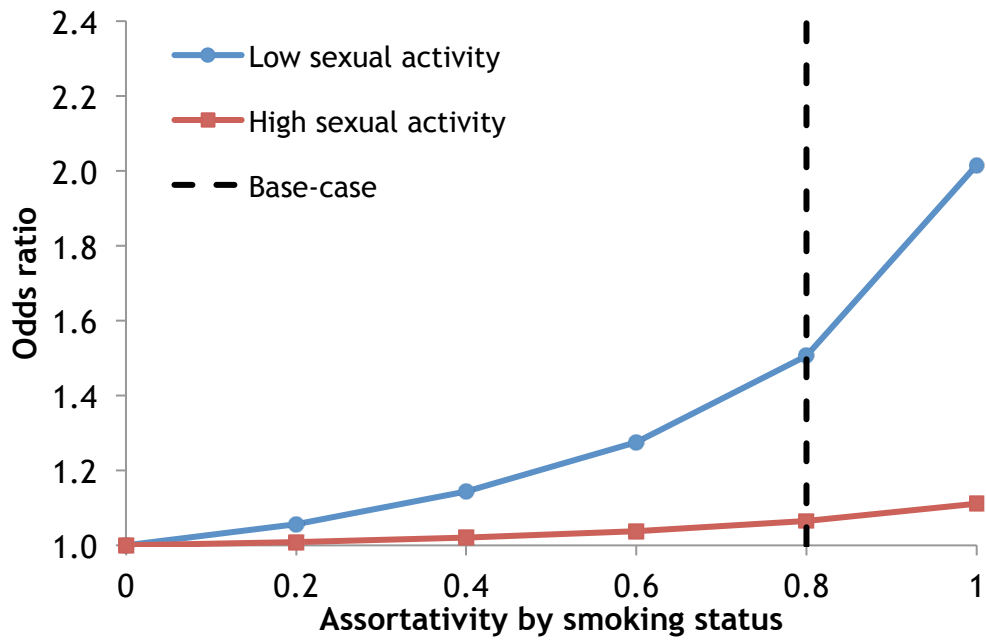
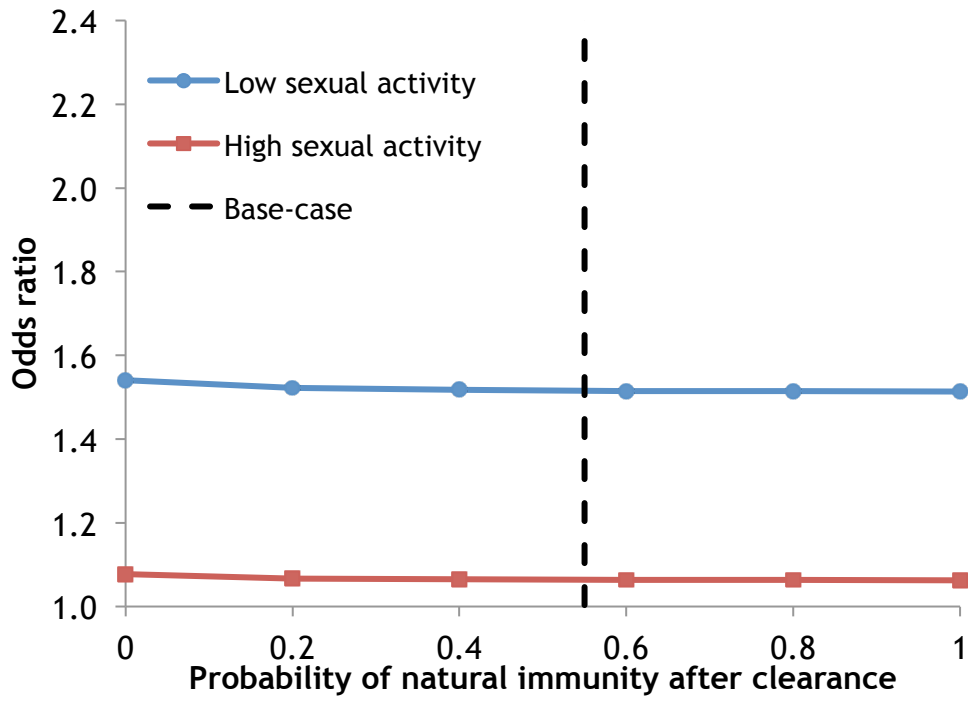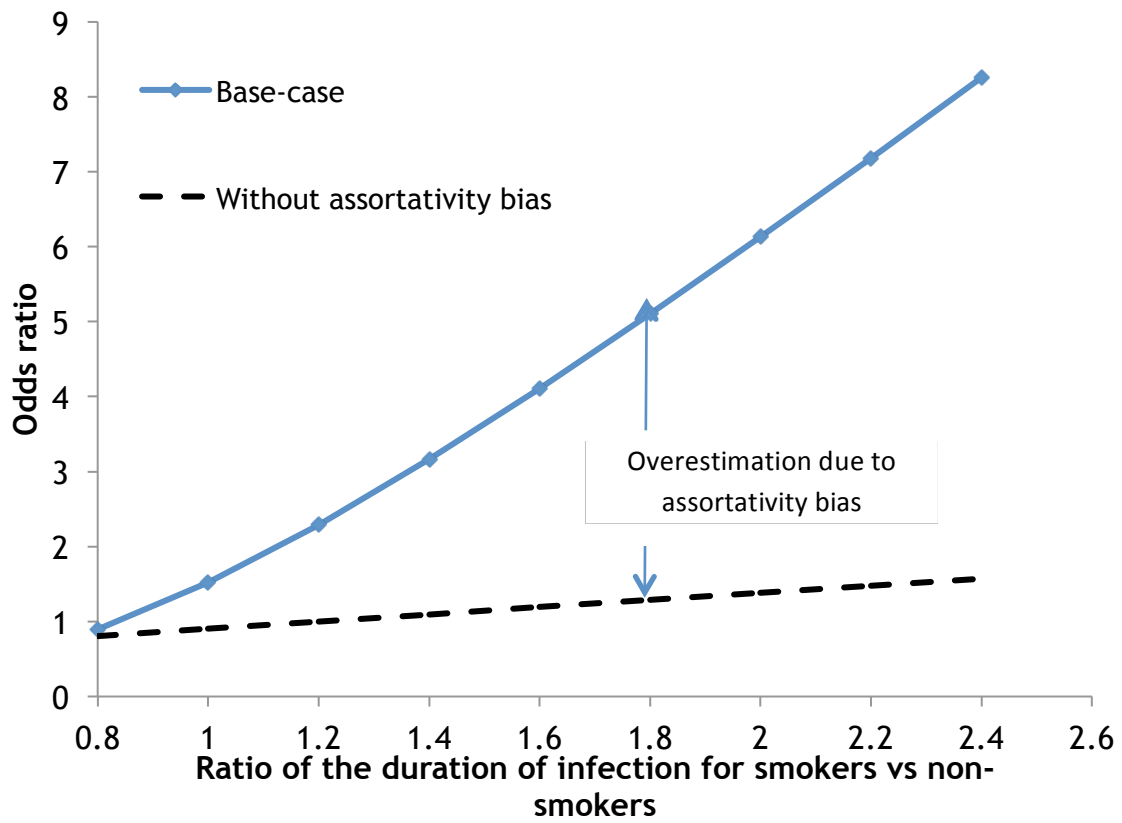**c)**
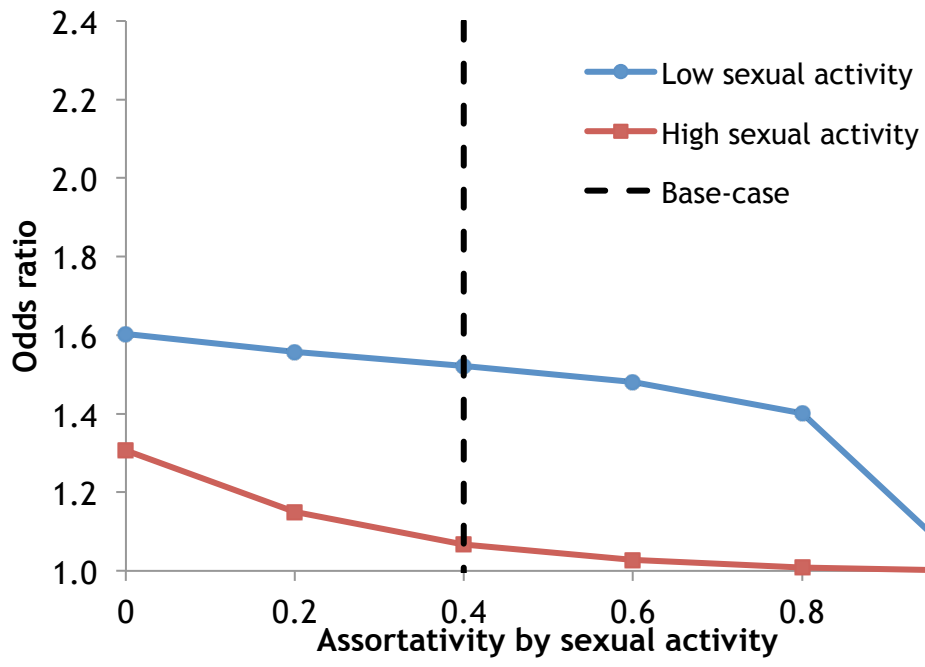
Figure 3

C)

Figure 4

REFERENCES

1.   Adimora AA, Schoenbach VJ. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *J Infect Dis* 2005;191 Suppl 1:S115-22.
2.   Fine P, Eames K, Heymann DL. "Herd immunity": a rough guide. *Clin Infect Dis* 2011;52(7):911-6.
3.   Schlecht NF, Franco EL, Rohan TE, et al. Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews. *J Epidemiol Biostat* 2001;6(5):393-407.
4.   Furber AS, Maheswaran R, Newell JN, et al. Is smoking tobacco an independent risk factor for HIV infection and progression to AIDS? A systemic review. *Sex Transm Infect* 2007;83(1):41-6.
5.   Doherty IA, Padian NS, Marlow C, et al. Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *J Infect Dis* 2005;191 Suppl 1:S42-54.
6.   Agrawal A, Heath AC, Grant JD, et al. Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behav Genet* 2006;36(4):553-66.
7.   Laumann EO, Youm Y. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. *Sex Transm Dis* 1999;26(5):250-61.
8.   Clark AE, Etile F. Don't give up on me baby: spousal correlation in smoking behaviour. *J Health Econ* 2006;25(5):958-78.
9.   Vaccarella S, Herrero R, Snijders PJ, et al. Smoking and human papillomavirus infection: pooled analysis of the International Agency for Research on Cancer HPV Prevalence Surveys. *Int J Epidemiol* 2008;37(3):536-46.
10.   Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, et al. Number of sexual partners and associations with initiation and intensity of substance use. *AIDS Behav* 2011;15(4):869-74.
11.   Boily MC, Anderson RM. Human immunodeficiency virus transmission and the role of other sexually transmitted diseases. Measures of association and study design. *Sex Transm Dis* 1996;23(4):312-32.
12.   Desai KN, Boily MC, Masse BR, et al. Simulation studies of phase III clinical trials to test the efficacy of a candidate HIV-1 vaccine. *Epidemiol Infect* 1999;123(1):65-88.
13.   White MT, Griffin JT, Drakeley CJ, et al. Heterogeneity in malaria exposure and vaccine response: implications for the interpretation of vaccine efficacy trials. *Malar J* 2010;9:82.
14.   Vickerman P, Foss A, Watts C. Using modeling to explore the degree to which a microbicide's sexually transmitted infection efficacy may contribute to the HIV effectiveness measured in phase 3 microbicide trials. *J Acquir Immune Defic Syndr* 2008;48(4):460-7.
15.   Wolf R, Freedman D. Cigarette smoking, sexually transmitted diseases, and HIV/AIDS. *Int J Dermatol* 2000;39(1):1-9.
16.   Gillison ML, Broutian T, Pickard RK, et al. Prevalence of oral HPV infection in the United States, 2009-2010. *JAMA* 2012;307(7):693-703.
17.   Beachler DC, Weber KM, Margolick JB, et al. Risk factors for oral HPV infection among a high prevalence population of HIV-positive and at-risk HIV-negative adults. *Cancer Epidemiol Biomarkers Prev* 2012;21(1):122-33.

18. Sellors JW, Mahony JB, Kaczorowski J, et al. Prevalence and predictors of human papillomavirus infection in women in Ontario, Canada. Survey of HPV in Ontario Women (SHOW) Group. *Cmaj* 2000;163(5):503-8.
19. Granath F, Giesecke J, Scalia-Tomba G, et al. Estimation of a preference matrix for women's choice of male sexual partner according to rate of partner change, using partner notification data. *Math Biosci* 1991;107(2):341-8.
20. Aral SO, Hughes JP, Stoner B, et al. Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *Am J Public Health* 1999;89(6):825-33.
21. Brisson M VdVN, Drolet M, Boily MC, Franco EL, Mayrand MH, et al. HPV-advise: technical appendix to J Natl Cancer Inst 2012; http://dx.doi.org/10.1093/jnci/djs395. http://www.marc-brisson.net/HPVadviseCEA.pdf. 2012.
22. Johnson HC, Elfstrom KM, Edmunds WJ. Inference of type-specific HPV transmissibility, progression and clearance rates: a mathematical modelling approach. *PLoS One* 2012;7(11):e49614.
23. Bogaards JA, Xiridou M, Coupe VM, et al. 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. *Am J Epidemiol* 2010;171(7):817-25.
24. Drolet M, Brisson M, Maunsell E, et al. The psychosocial impact of an abnormal cervical smear result. *Psychooncology* 2012;21(10):1071-81.
25. Manhart LE, Aral SO, Holmes KK, et al. Influence of study population on the identification of risk factors for sexually transmitted diseases using a case-control design: the example of gonorrhea. *Am J Epidemiol* 2004;160(4):393-402.
26. Liljeros F, Edling CR, Nunes Amaral LA. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes Infect* 2003;5(2):189-96.
27. Ley C, Bauer HM, Reingold A, et al. Determinants of genital human papillomavirus infection in young women. *J Natl Cancer Inst* 1991;83(14):997-1003.
28. Kjaer SK, van den Brule AJ, Bock JE, et al. 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 Epidemiol Biomarkers Prev* 1997;6(10):799-805.
29. Burk RD, Ho GY, Beardsley L, et al. Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women. *J Infect Dis* 1996;174(4):679-89.
30. Franco EL, Spence AR. Commentary: Smoking and human papillomavirus infection: the pursuit of credibility for an epidemiologic association. *Int J Epidemiol* 2008;37(3):547-8.
31. Wheeler CM, Hunt WC, Cuzick J, et al. A population-based study of human papillomavirus genotype prevalence in the United States: baseline measures prior to mass human papillomavirus vaccination. *Int J Cancer* 2013;132(1):198-207.
32. Chinsembu K. Sexually Transmitted Infections in Adolescents. *The Open Infectious Diseases Journal* 2009;3:107-17.
33. Johnson AM, Mercer CH, Erens B, et al. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet* 2001;358(9296):1835-42.
34. Chandra A, Mosher WD, Copen C, et al. Sexual behavior, sexual attraction, and sexual identity in the United States: data from the 2006-2008 National Survey of Family Growth. *Natl Health Stat Report* 2011(36):1-36.
35. Hariri S, Unger ER, Sternberg M, et al. Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006. *J Infect Dis* 2011;204(4):566-73.

36.   Cubbin C, Santelli J, Brindis CD, et al. Neighborhood context and sexual behaviors among adolescents: findings from the national longitudinal study of adolescent health. *Perspect Sex Reprod Health* 2005;37(3):125-34.
37.   Hiscock R, Bauld L, Amos A, et al. Socioeconomic status and smoking: a review. *Ann N Y Acad Sci* 2012;1248:107-23.
38.   Laumann EO GJ, Michael RT, Michaels S. *The social organization of sexuality*. Chicago: The University of Chicago Press; 1994.

39.   Burchell AN, Tellier PP, Hanley J, et al. Influence of partner's infection status on prevalent human papillomavirus among persons with a new sex partner. *Sex Transm Dis* 2010;37(1):34-40.
40.   Bosch FX, Castellsague X, Munoz N, et al. Male sexual behavior and human papillomavirus DNA: key risk factors for cervical cancer in Spain. *J Natl Cancer Inst* 1996;88(15):1060-7.
41.   Diez Roux AV, Aiello AE. Multilevel analysis of infectious diseases. *J Infect Dis* 2005;191 Suppl 1:S25-33.

Figure 1. Illustrated example of the *assortativity bias* in the association between smoking and a sexually transmitted infection. Directed arrows represent causal link and double-headed arrows represent statistical link. For simplicity, we assume that smoking status is the only factor determining partner selection and that smoking does not affect risk and duration of infection.

Figure 2. Impact of behavioral parameters on the *assortativity bias*. Univariate sensitivity analysis of the odds ratios of prevalence between smokers and non-smokers with one parameter varying: A) Proportion of smokers that are highly sexually active, B) Assortativity by smoking status, and C) Assortativity by sexual activity. For panel A), the proportion of non-

smokers that are highly sexually active is fixed at its base case value. Hence, increasing the parameter in A) increases the strength of the association between smoking and sexual activity.

Figure 3. Impact of biological parameters on the *assortativity bias*. Univariate sensitivity analysis of the odds ratios of infection between smokers and non-smokers varying: A) Probability of transmission per partnership, B) Duration of infection, and C) Probability of developing natural immunity after clearance of infection.

Figure 4. Impact of a biological effect of smoking on the duration of infection. Univariate sensitivity analysis of the odds ratio (OR) of prevalence between smokers and non-smokers varying: Ratio of smokers' versus non-smokers' duration of infection. Two scenarios are shown both with base-case parameters except for the assortativity by smoking status: the blue curve is a scenario with assortativity parameter of 0.8 as in the base case and the dashed curve is a scenario without assortativity (parameter of 0). Hence, the difference in height between the two curves measures the magnitude of the overestimation due to the *assortativity bias*.