

# Assortative mixing as a source of bias in epidemiological studies of sexually transmitted infections: the case of smoking and human papillomavirus

P. LEMIEUX-MELLOUKI<sup>1,2</sup>, M. DROLET<sup>1,2</sup>, J. BRISSON<sup>1,2</sup>, E. L. FRANCO<sup>3</sup>,  
M.-C. BOILY<sup>4</sup>, I. BAUSSANO<sup>5</sup> AND M. BRISSON<sup>1,2,4\*</sup>

<sup>1</sup> *CHU de Québec Research Center, Québec, Canada*

<sup>2</sup> *Department of Social and Preventive Medicine, Laval University, Québec, Canada*

<sup>3</sup> *Division of Cancer Epidemiology, McGill University, Montreal, Canada*

<sup>4</sup> *Department of Infectious Disease Epidemiology, Imperial College, London, UK*

<sup>5</sup> *International Agency for Research on Cancer, Lyon, France*

*Received 14 July 2015; Final revision 23 September 2015; Accepted 28 October 2015;  
first published online 20 November 2015*

## SUMMARY

For studies examining risk factors of sexually transmitted infections (STIs), confounding can stem from characteristics of partners of study subjects, and persist after adjustment for the subjects' individual-level characteristics. Two conditions that can result in confounding by the subjects' partners are: (C1) partner choice is assortative by the risk factor examined and, (C2) sexual activity is associated with the risk factor. 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–C2 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 stronger association of smoking with sexual activity. Adjustment for characteristics of partners is needed to mitigate assortativity bias.

**Key words:** Epidemiology, human papilloma virus (HPV), mathematical modelling sexually transmitted infections (STIs).

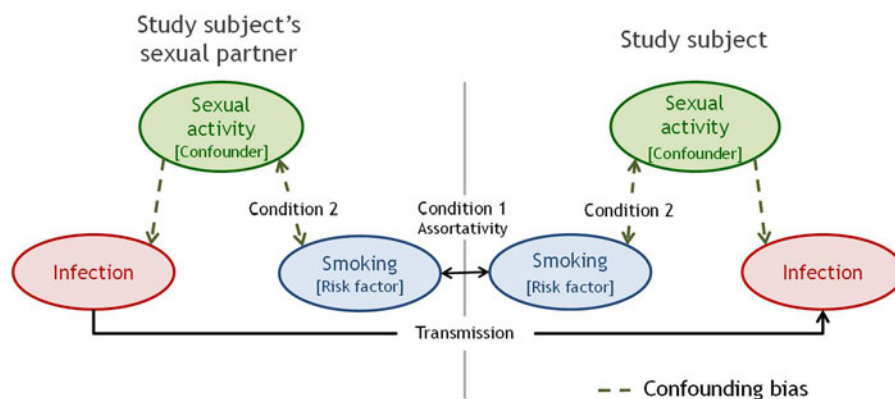
## INTRODUCTION

Contact networks of individuals affect their exposure to infectious contacts and are therefore a crucial determinant of infection risk [1]. Thus, an individual's risk of infection not only depends on individual-level risk factors,

such as gender, but on network-level risk factors. A classic example of this particularity of infectious diseases is herd protection: 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 observational studies that examine risk factors of sexually transmitted infections (STIs), control

\* Author for correspondence: Dr M. Brisson, Centre de recherche du CHU de Québec, Axe Santé des populations et pratiques optimales en santé, 1050 Chemin Sainte-Foy, Québec, Canada, G1S 4L8.  
(Email: Marc.Brisson@crchudequebec.ulaval.ca)



**Fig. 1.** Illustrated example of the assortativity bias in the association between smoking and a sexually transmitted infection. Direction 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.

of confounding most often follows the traditional non-transmissible disease approach of controlling for individual-level risk factors (such as the subject's sexual activity), with little attention to control for network-level risk factors such as sexual activity of individuals in the subject's network. In doing so, authors usually acknowledge the possibility of misclassification of sexual activity (e.g. number of partners), due to misreporting by study subjects, which can cause residual confounding if sexual activity is associated with the risk factors [3, 4]. What is less acknowledged in these studies is that, even if the sexual activity of the study subjects was perfectly measured and controlled for, the sexual activity of individuals in the subject's network can also confound the biological effect of a risk factor on acquisition of infection. We term one particular form of such confounding '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 behaviour (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 confounding is likely to remain even after perfectly controlling the effect measure for individual-level confounders.

To illustrate assortativity bias, we consider a simplified example where the effect of smoking on the risk of a STI is examined. Thus, smoking is the

exposure variable and the occurrence of a STI is the outcome. 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 smokers have a higher average level of sexual activity (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 relationship is confounded by sexual activity of the subjects (Fig. 1, right panel, dotted arrows) and subjects' partners (Fig. 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). Of note, 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. Hence, the assortativity bias ultimately reflects differences in sexual network of smokers compared to non-smokers, and not only differences in sexual activity of subjects' current partners. Even if we assume a biological effect of smoking on STI,

the assortativity bias would still occur resulting in an overestimation of this effect. Given that C1 and C2 are met for many common risk factors of STIs, including age, race and socioeconomic status (SES), and the assortativity bias affects measures of STI acquisition (e.g. incidence rate and prevalence ratios), the *bias* is likely present in many prospective and cross-sectional epidemiological STI studies.

Mathematical modelling has been used to understand potential biases in epidemiological studies of STIs [11–15]. With modelling, an artificial world is created where transmission and natural history of disease can be simulated based on model inputs which are either assumed or fitted to empirically observed data. Epidemiological studies can be nested within the model to examine potential biases under different assumptions regarding behaviour, 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 [16]. As smoking is a modifiable behaviour, 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, 17–19], and these associations have been shown to follow monotone dose-response relationships [9]. However, the possibility of assortativity bias was not addressed in these studies.

The specific objectives of this paper are to use mathematical modelling: (1) to illustrate and describe the assortativity bias, using as an example the association between smoking and HPV infection, and (2) to examine the sensitivity of the assortativity bias to biological and behavioural parameters, for generalization of results.

## METHODS

### Mathematical model

We developed a deterministic transmission-dynamic model of HPV infection (see Supplementary material for a list of the model's equations). The modelled population is heterosexual, open and stable. For the base-case scenario, we modelled HPV16 infection, which is the most prevalent and oncogenic type. The simulated population is stratified for the two behavioural aspects from which 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 modelled population, representing the years of higher sexually active life (ages 15–44 years). Sexual mixing depends on an individual's smoking status and sexual activity. For each of these behavioural factors, we allowed mixing to vary from random to completely assortative (smokers only form partnerships with smokers). We included assortativity according to sexual activity as it is a key feature of sexual networks [20, 21]. 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 has no biological effect (smoking does not increase HPV transmission probabilities or duration), to test whether observed associations between smoking and HPV infection can be explained by the assortativity bias.

### Parameterization

Model parameter values and references are presented in Table 1. We used biological parameter values estimated in prior modelling work [22–24], and estimated the proportion of smokers in each sexual activity group from an epidemiological study [25]. Although studies suggest that sexual mixing by level of sexual activity and smoking status is assortative [6, 8, 26, 27], no empirical estimates of assortativity are available in the literature. In our base case, we assumed assortativity parameter values for smoking status and for sexual activity to be 0.8 and 0.4, respectively (0.0 = random, 1.0 = complete assortativity), using equations presented in the Supplementary material. We performed extensive sensitivity analysis on mixing parameters 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 simulated population. HPV prevalence was estimated at endemic equilibrium without HPV vaccination with perfect sensitivity and specificity. We used odds ratios (ORs) of HPV infection (positivity) in smokers compared to non-smokers as the measure of association.

Table 1. *Base-case model parameter values*

Population parameters	Symbol	Base case	Reference
Natural mortality rate (per year)	$\mu$	1/30	Assumption
Probability of transmission per partnership	$\pi$	75%	[22–24]
Average duration of infection (years)	$D = 1/\sigma$	1.2	[22]
Probability of developing natural immunity after clearance of infection	$\omega$	30%	[22]
Proportion of smokers in the population	$\sigma$	20%	[25]
Proportion of individuals in the high activity group	$a_S$	1.2%	[25]
Risk ratio of being highly sexually active if smoker compared to non-smoker	$a_{NS}$	7.5	[25]
Rate of new sexual partners per year in sexual activity group $i$	$n_i$	$n_1 = 1.2$ $n_2 = 7.0$	[25]
Sexual assortativity parameter	$\varepsilon_1$	0.4	Assumption
Smoking assortativity parameter	$\varepsilon_2$	0.8	Assumption
Parameter of correlation between the two types of assortativity	$C$	1	Assumption

The overall adjusted ORs were calculated as the weighted average of the stratum-specific ORs of the two sexual activity groups using, as weights, the proportion of the population in each sexual activity group (see Supplementary material for the equations used to compute ORs). In the simulated study, adjusted or stratum-specific ORs different from 1.00 can only be due to assortativity bias, because there is no biological effect of smoking in our model, 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. Hence, the simulation reproduces the conduct of a perfect study with no other biases, be it misclassification or confounding, other than the assortativity bias. Because the simulated population is at equilibrium and the duration of infection is assumed to be unaffected by smoking, the incidence rate ratios from nested longitudinal studies in our model have the same numerical value to the ORs from cross-sectional studies: a result given by the formula:

$$\frac{\text{prevalence}}{1 - \text{prevalence}} = \text{incidence} \times \text{duration}.$$

### Sensitivity analyses

We varied the key biological/behavioural 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 group. Finally, we estimated the potential impact of the assortativity bias on adjusted ORs assuming different magnitudes of a true biological effect of smoking on infection.

Table 2. *Odds ratios of HPV infection between smokers and non-smokers in modelled study subjects*

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

## RESULTS

### Base case

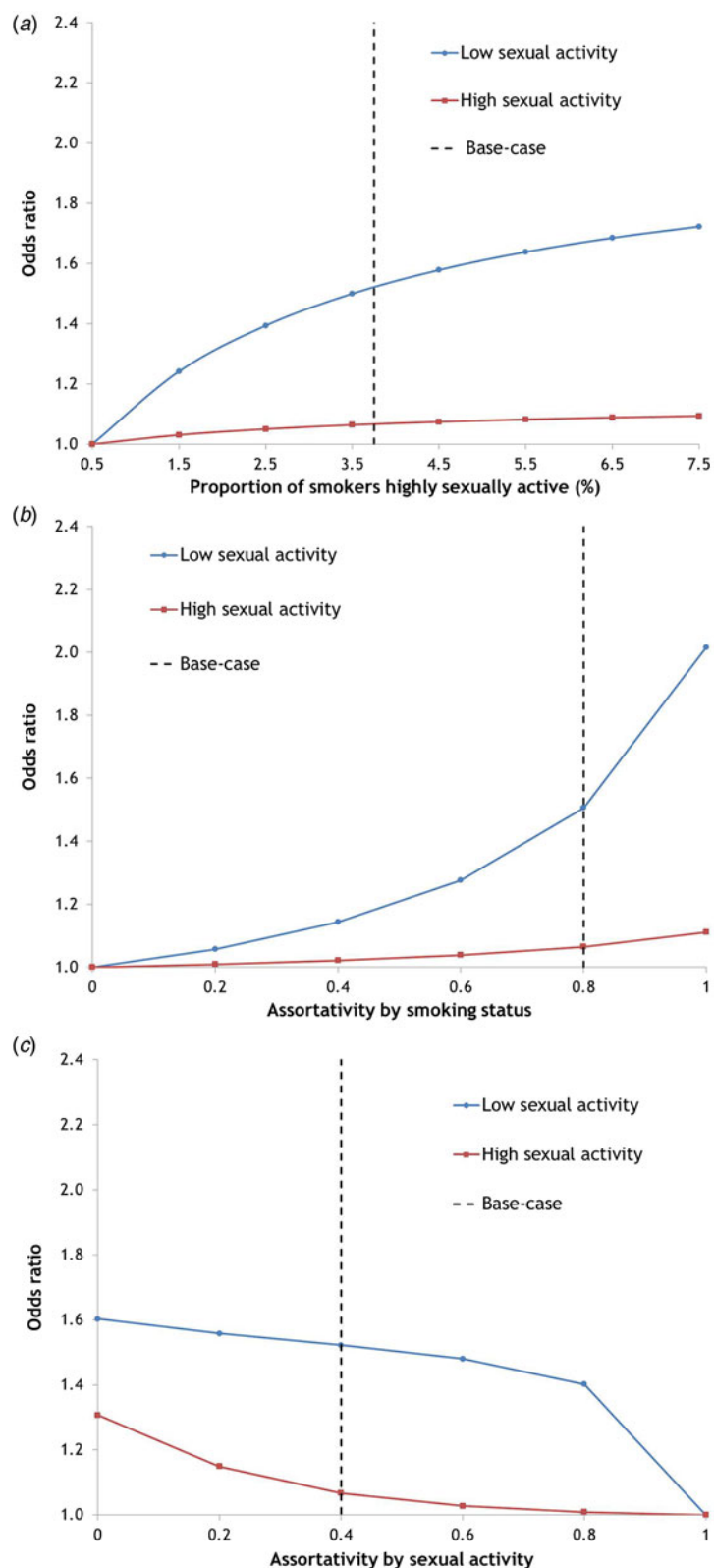
Table 2 shows the base-case model predictions of the ORs of HPV infection in smokers compared to non-smokers. We estimated crude and adjusted ORs of 1.64 and 1.51, respectively (Table 2). Given that, in our model, smoking has no causal effect on HPV and we can perfectly control for sexual activity of subjects (no residual confounding), assortativity bias is the only possible cause of adjusted OR > 1.00.

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

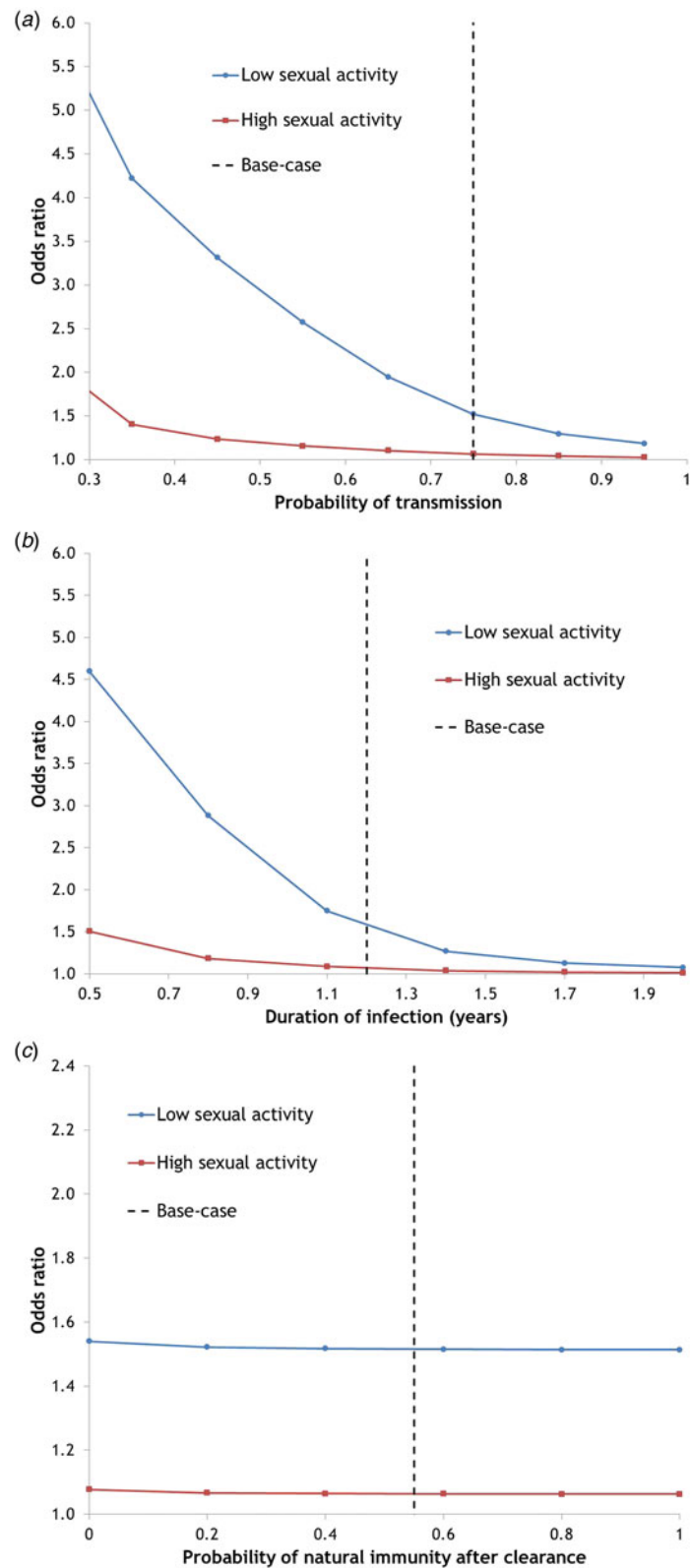
### Impact of behavioural factors

#### *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 in study subjects increases (Fig. 2a). This is because a stronger association causes



**Fig. 2.** Impact of behavioural 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.



**Fig. 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.



greater confounding by sexual activity (increased impact of C2, Fig. 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 (Fig. 2*b*). When smoking assortativity is stronger, the imbalance 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 C1, Fig. 1).

#### *Assortativity by sexual activity*

The ORs of HPV infection decrease with greater assortativity between individuals of the same sexual activity group (Fig. 2*c*). As assortativity by sexual activity increases, the sexual activity of the study subject becomes a better proxy of his/her partners' sexual activity. When mixing by sexual activity is completely assortative, subjects will have partners belonging to their own sexual activity group, irrespective of smoking status. Therefore, there will be no bias after adjustment for sexual activity.

### **Impact of biological factors**

#### *Transmission probability and duration of infectiousness*

The model shows that the magnitude of the assortativity bias is highly sensitive to the transmission probability or duration of infection (Fig. 3*a, b*). The ORs of HPV infection in smokers compared to 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 ( $R_0$ ) 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.

#### *Natural immunity*

The probability of developing natural immunity has little impact on the magnitude of the assortativity bias (Fig. 3*c*). Lower natural immunity has the same relative impact on HPV prevalence in both smokers and non-smokers.

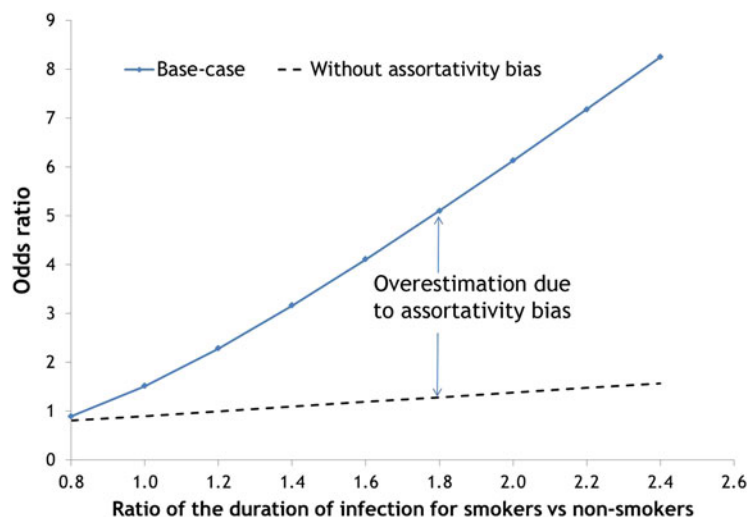
### **Assortativity bias assumes a true biological effect of smoking**

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 rises steeply as the biological effect of smoking increases. The OR is also overestimated when smoking affects the transmission probability, or the probability of developing natural immunity (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 modelling, we showed that adjustment for the subjects' individual-level sexual activity is insufficient to attribute the association between smoking and HPV to a biological effect 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 STIs such as age, race/ethnicity and SES. Hence, the assortativity bias is likely present in many epidemiological studies examining risk factors of STIs.

Our modelling analysis suggests that the assortativity bias could produce ORs of the magnitude seen in empirical studies on HPV if assortativity by smoking status is high. In a recent large-scale study, the adjusted ORs of HPV infection in smokers compared to non-smokers was 1.4 (95% confidence interval 1.2–1.7) [9], and most other studies have found ORs higher than 1.0 [19, 28–30]. The association between smoking and HPV infection is supported by traditional criteria of causality such as dose-response. However, the assortativity bias can produce 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 empirically based studies for many other STIs [4, 17, 18]. We also showed that the size of the assortativity bias should be more important for STIs such as HIV, which have low  $R_0$ .



**Fig. 4.** Impact of a biological effect of smoking on the duration of infection. Univariate sensitivity analysis of the odds ratio of prevalence between smokers and non-smokers varying: ratio of smokers' vs. 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.

Our results should not be interpreted as evidence that smoking is not a cause of HPV infection. Smoking 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 [31]. 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 substantially between studies due to differences in the behaviour of participants (differences in the magnitude of C1–C2).

The assortativity bias could affect many risk factors other 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 of subjects [26, 32]. It is suggested that this is due to a biological cause (e.g. cervical ectopy makes young women vulnerable to STIs) [33]. Yet, sexual mixing is highly assortative with respect to age [21, 34] (C1), and younger adults are more sexually active [34, 35] (C2), and hence an age–STI association can be partly due to the assortativity bias. For other risk factors, complete assortativity between individuals with the risk factor can hold automatically and cause assortativity bias in prevalence studies. For example, in a cross-sectional study examining

HPV as a risk factor of another STI, a subject infected with HPV will have a previous/current partner also infected with HPV. However, subjects' partners infected with HPV will have greater sexual activity on average and thus higher risk of other STIs. Hence, HPV can be identified 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 modelling to perfectly control a fictive population, allowing us to explore the theoretical basis for the bias and the relationship between the bias and behavioural and biological parameters. However, the main limitation of our model is that many sources of heterogeneity (sexual activity, smoking intensity) were not included and we assumed independence between mixing by sexual activity and by smoking status. Greater heterogeneity in sexual activity would require specifying in C2 that the association between sexual activity and smoking is monotonic, which seems to be the case [25]. Furthermore, 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 [36, 37], and sexual mixing is assortative by SES [37], which indirectly produces assortativity with respect to smoking. In this case, the 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, one should not use the precise model OR estimates as being representative of reality.

To correct for the assortativity bias in studies examining risk factors of STIs, one must control for systematic differences in exposure to infection that can occur between individuals with and without a given risk factor (e.g. smoking). To control for differences in exposure to infection, studies have restricted their population to individuals known to have been exposed to infection. For instance, studies examining risk factors of HIV transmission have used populations of serodiscordant couples [38, 39], where the uninfected partner is known to be exposed. However, such studies are rarely performed for other STIs, as they are costly and methodologically challenging (difficult to adequately condition on exposure to infection). Randomized trials on the other hand would suffer from the assortativity bias if the treatment is a cause of assortativity and if subjects can acquire sexual partnership after the randomization. If sexual partnerships are stable from the randomization until the end of follow-up, there remains the difficulty of interpreting the measure of effect because of the absence of conditioning on exposure to infection. Furthermore, not all causal factors can be investigated in randomized trials (e.g. smoking, age) and only one factor can be examined per trial. Hence, most studies examining STI risk factors are based on cross-sectional or prospective data, where infection status and risk factors are assessed without specific data on exposure to infection. For such studies, the characteristics of the study subjects' sexual partners should be used to reduce the assortativity bias. Taking the example of smoking, the smoking status of new sexual partners of study subjects should be assessed in prospective studies to control for the higher chance of smokers having partners who are smokers. In addition, information on past partners would also be needed, with a recall window depending on duration of infection. Given that many risk factors are investigated at once in empirical studies, it is also necessary to have simultaneous adjustments for the key risk factors being investigated (e.g. age, race/ethnicity, SES) at the subject- and partner-level.

In conclusion, assortative sexual mixing by smoking status can cause bias in studies assessing the biological effect of smoking on HPV acquisition. 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.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268815002915>.

## ACKNOWLEDGEMENTS

This work was supported by the Canadian Institutes of Health Research (GSD- 130 809 to P.L.M.), and the Canada Research Chair programme (to M.B.). The funding sources had no involvement in the study design and conduct of the study; and preparation, review, or approval of the manuscript.

## DECLARATION OF INTEREST

Over the past 3 years, M.B. has received an unrestricted grant from Merck (herpes zoster - none are ongoing). Although unrelated to the content and message of the manuscript, M.D. has consulted for GlaxoSmithKline (herpes zoster vaccine); E.L.F. receives occasional consultancy fees from companies involved with HPV diagnostics (Roche, Qiagen, Gen-Probe) and has also served as an occasional paid consultant to companies involved with HPV vaccines (Merck, GSK).

## REFERENCES

1. **Adimora AA, Schoenbach VJ.** Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *Journal of Infectious Diseases* 2005; **191** (Suppl. 1): S115–122.
2. **Fine P, Eames K, Heymann DL.** 'Herd immunity': a rough guide. *Clinical Infectious Diseases* 2011; **52**: 911–916.
3. **Schlecht NF, et al.** Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews. *Journal of Epidemiology and Biostatistics* 2001; **6**: 393–407.
4. **Furber AS, et al.** Is smoking tobacco an independent risk factor for HIV infection and progression to AIDS? A systemic review. *Sexually Transmitted Infections* 2007; **83**: 41–46.
5. **Doherty IA, et al.** Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *Journal of Infectious Diseases* 2005; **191** (Suppl. 1): S42–54.
6. **Agrawal A, et al.** Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behavior Genetics* 2006; **36**: 553–566.
7. **Laumann EO, Youm Y.** Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. *Sexually Transmitted Diseases* 1999; **26**: 250–261.

8. Clark AE, Etile F. Don't give up on me baby: spousal correlation in smoking behaviour. *Journal of Health Economics* 2006; **25**: 958–978.
9. Vaccarella S, et al. Smoking and human papillomavirus infection: pooled analysis of the International Agency for Research on Cancer HPV Prevalence Surveys. *International Journal of Epidemiology* 2008; **37**: 536–546.
10. Cavazos-Rehg PA, et al. Number of sexual partners and associations with initiation and intensity of substance use. *AIDS and Behavior* 2011; **15**: 869–874.
11. Boily MC, Anderson RM. Human immunodeficiency virus transmission and the role of other sexually transmitted diseases. Measures of association and study design. *Sexually Transmitted Diseases* 1996; **23**: 312–332.
12. Desai KN, et al. Simulation studies of phase III clinical trials to test the efficacy of a candidate HIV-1 vaccine. *Epidemiology and Infection* 1999; **123**: 65–88.
13. White MT, et al. Heterogeneity in malaria exposure and vaccine response: implications for the interpretation of vaccine efficacy trials. *Malaria Journal* 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. *Journal of Acquired Immune Deficiency Syndromes* 2008; **48**: 460–467.
15. Koopman JS, et al. Assessing risk factors for transmission of infection. *American Journal of Epidemiology* 1991; **133**: 1199–1209.
16. Wolf R, Freedman D. Cigarette smoking, sexually transmitted diseases, and HIV/AIDS. *International Journal of Dermatology* 2000; **39**: 1–9.
17. Gillison ML, et al. Prevalence of oral HPV infection in the United States, 2009–2010. *Journal of the American Medical Association* 2012; **307**: 693–703.
18. Beachler DC, et al. Risk factors for oral HPV infection among a high prevalence population of HIV-positive and at-risk HIV-negative adults. *Cancer Epidemiology, Biomarkers & Prevention* 2012; **21**: 122–133.
19. Sellors JW, et al. Prevalence and predictors of human papillomavirus infection in women in Ontario, Canada. Survey of HPV in Ontario Women (SHOW) Group. *Canadian Medical Association Journal* 2000; **163**: 503–508.
20. Granath F, 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. *Mathematical Biosciences* 1991; **107**: 341–348.
21. Aral SO, et al. Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *American Journal of Public Health* 1999; **89**: 825–833.
22. Van de Velde N, et al. Population-level impact of the bivalent, quadrivalent, and nonavalent human papillomavirus vaccines: a model-based analysis. *Journal of the National Cancer Institute* 2012; **104**: 1712–1723.
23. Johnson HC, Elfstrom KM, Edmunds WJ. Inference of type-specific HPV transmissibility, progression and clearance rates: a mathematical modelling approach. *PLoS ONE* 2012; **7**: e49614.
24. Bogaards JA, 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. *American Journal of Epidemiology* 2010; **171**: 817–825.
25. Drolet M, et al. The psychosocial impact of an abnormal cervical smear result. *Psychooncology* 2012; **21**: 1071–1081.
26. Manhart LE, et al. Influence of study population on the identification of risk factors for sexually transmitted diseases using a case-control design: the example of gonorrhoea. *American Journal of Epidemiology* 2004; **160**: 393–402.
27. Liljeros F, Edling CR, Nunes Amaral LA. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and Infection* 2003; **5**: 189–196.
28. Ley C, et al. Determinants of genital human papillomavirus infection in young women. *Journal of the National Cancer Institute* 1991; **83**: 997–1003.
29. Kjaer SK, 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 Epidemiology, Biomarkers and Prevention* 1997; **6**: 799–805.
30. Burk RD, et al. Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women. *Journal of Infectious Diseases* 1996; **174**: 679–689.
31. Franco EL, Spence AR. Commentary: Smoking and human papillomavirus infection: the pursuit of credibility for an epidemiologic association. *International Journal of Epidemiology* 2008; **37**: 547–548.
32. Wheeler CM, et al. A population-based study of human papillomavirus genotype prevalence in the United States: baseline measures prior to mass human papillomavirus vaccination. *International Journal of Cancer* 2013; **132**: 198–207.
33. Chinsambu K. Sexually transmitted infections in adolescents. *Open Infectious Diseases Journal* 2009; **3**: 107–117.
34. Johnson AM, et al. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet* 2001; **358**: 1835–1842.
35. Chandra A, et al. Sexual behavior, sexual attraction, and sexual identity in the United States: data from the 2006–2008 National Survey of Family Growth. *National Health Statistics Report* 2011:1–36.
36. Hiscock R, et al. Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences* 2012; **1248**: 107–123.
37. Laumann EO GJ, Michael RT, Michaels S. *The Social Organization of Sexuality*. Chicago: The University of Chicago Press, 1994.
38. Cohen MS, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine* 2011; **365**: 493–505.
39. Zhang YJ, et al. HIV transmission and related risk factors among serodiscordant couples in Liuzhou, China. *Journal of Medical Virology* 2015; **87**: 553–556.