Modelling changes in sexual behavior and multi-site transmission of HPV to assess their impact on past and future trends of HPV infections and diseases

Thèse

Philippe Lemieux-Mellouki

Doctorat en épidémiologie
Philosophiæ doctor (Ph. D.)

Québec, Canada

Modelling changes in sexual behavior and multi-site transmission of HPV to assess their impact on past and future trends of HPV infections and diseases

Thèse

Philippe Lemieux-Mellouki

Sous la direction de:

Marc Brisson, directeur de recherche

Résumé

Introduction: Lors de l'élaboration des modèles de dynamiques de transmission du virus du papillome humain (VPH), différentes hypothèses sont émises pour simplifier les aspects complexes et peu compris de la transmission. Dans cette thèse, on s'intéresse à deux limites de ces modèles: 1) la transmission du VPH entre différents sites n'est pas modélisée, 2) les changements dans l'activité sexuelle à travers le temps sont ignorés. Les changements dans l'activité sexuelle au cours du dernier siècle pourraient être l'une des causes de l'augmentation des cancers de l'oropharynx des quatre dernières décennies. Cependant, notre compréhension des tendances des cancers reliés au VPH est entravée par le manque de connaissance des patrons de contacts sexuels. Un tel type de patron, l'assortativité, pourrait produire de la confusion dans les études épidémiologiques portant sur les facteurs de risques du VPH. Les objectifs de cette thèse sont 1) d'évaluer l'impact d'inclure la transmission multi-site du VPH dans un modèle sur les prédictions d'efficacité populationnelle du vaccin VPH, 2) d'examiner comment les changements dans l'activité sexuelle ont pu influencer les tendances des cancers de l'oropharynx (CO) et du col de l'utérus (CU) depuis les années 1970, et de prédire les tendances futures de ces cancers, 3) déterminer les conditions sous lesquelles l'assortativité pourrait causer un biais dans les études évaluant la causalité des facteurs de risques des infections transmissibles sexuellement et quantifier la magnitude de ce biais.

Méthodes : Pour le premier objectif, nous avons développé un modèle de transmission du VPH multi-site (sites génital et extra-génital) et uni-site (site génital). Avec ces modèles, nous avons estimé la réduction relative de la prévalence du VPH au site génital après la vaccination (RR_{prev}). Nous avons considéré deux types d'immunité naturelle : site-spécifique et systémique. Pour le deuxième objectif, nous avons développé un modèle mathématique individus-centré simulant la transmission du VPH16, la progression du VPH16 vers le cancer (CO et CU) et les comportements sexuels des américains nés entre 1850 et 1999. Nous avons calibré ce modèle et réalisé des simulations de l'incidence de CO et CU entre 1915 et 2045. Pour le troisième objectif, nous avons développé un modèle de transmission du VPH avec stratification pour deux niveaux d'activité sexuelle (élevé et faible) et de tabagisme (fumeur et non-fumeur). On a supposé dans ce modèle que le tabagisme n'était pas une cause biologique d'infection au VPH, et que le choix du partenaire sexuel est assortatif au statut fumeur. Nous avons simulé une étude fictive dans laquelle nous avons estimé le rapport de cotes (RC) de la prévalence d'infection VPH entre fumeurs et non-fumeurs. La magnitude du biais se mesurait

par l'écart entre le RC et la valeur nulle (1,00).

Résultats : Pour l'objectif 1, le modèle multi-site prédisait un RR_{prev} supérieur à celui estimé par le modèle uni-site quand l'immunité était site-spécifique, et inférieur quand l'immunité était systémique. La magnitude de la transmission entre les sites génital et extra-génital était un facteur expliquant la variance du RR_{prev}. Pour l'objectif 2, notre modèle a prédit, en absence de dépistage du CU, une augmentation importante de l'incidence de CU d'au moins 120% entre 1975 et 2015. Pour le cancer de l'oropharynx relié au VPH16, le modèle a prédit une augmentation d'au moins 310% entre 1985 et 2015, et d'environ 50% entre 2015 et 2045. Pour l'objectif 3, nous avons obtenu un RC de 1,51 après ajustement parfait du niveau d'activité sexuel des sujets de l'étude. Une plus grande assortativité dans le choix du partenaire sexuel causait une augmentation de la magnitude du biais.

Conclusion : Étant donné les connaissances actuelles dans le domaine du VPH, il semble peu probable que les prédictions d'efficacité populationnelle de la vaccination contre le VPH faites avec un modèle uni-site soient significativement biaisées. Cependant, l'utilisation d'un modèle multi-site nous a permis de reproduire les changements de comportements sexuels, d'expliquer les tendances de CO observées depuis les années 1980 et de prédire une augmentation future dans l'incidence de CO. Toutefois, cette augmentation future ne pourra pas être prévenue par la vaccination car elle touche surtout des hommes nés avant 1990, qui n'ont pas été vaccinés. Finalement, le rôle du tabagisme dans l'acquisition du VPH demeure incertain dû au biais d'assortativité que nous avons identifié .

Abstract

Introduction: Dynamic models of HPV transmission have been the main tool to estimate the effectiveness and cost-effectiveness of different vaccination strategies. We focus on two limitations of these models: transmission of HPV across different sites (e.g., oral and anal) and changes in sexual behavior across birth cohorts are not modelled. Firstly, including multi-site transmission of HPV could affect estimates of effectiveness against genital diseases. Secondly, changes in sexual behavior throughout the last century could be causally related to the increase in oropharyngeal and anal cancers in the past four decades. Understanding these trends is hindered by our poor knowledge of sexual mixing patterns. One such pattern, assortative mixing, could cause confounding in epidemiological studies of HPV risk factors. The objectives of this thesis are to 1) Assess the impact of including multi-site transmission to current uni-site HPV models on predictions of the population-level effectiveness of HPV vaccination. 2) Examine how changes in specific aspects of sexual behavior such as mixing, rates of new partner acquisition, sexual practices, age of sexual debut, may have impacted trends in HPV-related oropharyngeal and cervical cancer incidence since the 1970s, and predict future trends in these cancers. 3) Determine conditions under which assortative mixing could cause bias in studies examining the causal role of risk factors of STIs and quantify the magnitude of this potential bias.

Methods: For the first objective, we developed a multi-site (genital and extragenital sites) and a uni-site (genital site) model of HPV transmission. We estimated the reduction in genital HPV prevalence at equilibrium post-vaccination and compared the estimates of the two models. We considered two types of natural immunity: local (i.e., protects against subsequent infection at the same site) and systemic. For the second objective, we developed an individual-based model of HPV transmission at the genital and oral sites. We reproduced in this model the changes in sexual behavior from 1900 to 2015 in the US population, and simulated the incidence of HPV16-related oropharyngeal and cervical cancers between 1915 and 2045. We performed these simulations according to different scenarios regarding the practice of oral sex, the inclusion of oral infections, and the reporting bias in the number of sexual partners in surveys. Results of the simulations were compared with empirical data on HPV-related cancers. For the third objective, we developed a model of HPV transmission with stratifications for two levels of sexual activity and of smoking habits (smokers and non-smokers). In our simulation, smoking was not a biological cause of HPV infection. We then estimated the odds ratio of prevalent HPV infection between smokers and non-smokers. Deviation from the null value could

only be due to a bias we termed *assortativity bias*.

Results: For objective 1, multi-site model predicted higher vaccination effectiveness when natural immunity was local, and lower or equal effectiveness when natural immunity was systemic. Three important factors were identified to increase effectiveness predicted with the multi-site model: 1) higher proportion of genital infections caused by an extragenital infection, 2) lower proportion of extragenital infections caused by a genital infection, 3) higher proportion of susceptibles to extragenital infection. For objective 2, we predicted a sharp increase (IRR=[220%-380%]) between 1975 and 2015 in the simulations of cervical cancer incidence without cervical screening. The increase was lowest when assuming women under-report their number of sexual partners in surveys. In simulations of oropharyngeal cancer incidence, including past changes in the practice of oral sex produced a sharp increase between 1985 and 2015 similar to the observed US trends. Future incidence of oropharyngeal cancer was predicted to increase by 50% between 2015 and 2045. For objective 3, we obtained an OR of 1.51 after perfect adjustment for subjects' level of sexual activity. The non-biased OR for this simulation was 1.00. The magnitude of the bias, as measured by the deviation from the null value, increased with stronger association between sexual activity and smoking habits and with greater degree of assortativity with respect to smoking habits.

Conclusion: The assumptions of natural immunity being local and of extragenital infections being an important reservoir for genital infections are not currently supported by evidences. Hence, a significant bias in estimates of vaccination effectiveness against genital infections and diseases from the use of uni-site models appears unlikely. However, using a multi-site model and including the practice of oral sex is necessary to reproduce trends in oropharyngeal cancer since the 1980s. Furthermore, the increase in oropharyngeal cancer is predicted to continue over the next three decades, affecting mainly unvaccinated men born before 1990. In addition, we predict that cervical screening prevented a sharp increase in cervical cancer in the past decades. Finally, the role of smoking in the acquisition and duration of HPV infection remains uncertain due to biases such as the *assortativity bias*.

Contents

Liste of Publications

Lemieux-Mellouki P, Drolet M, Jit M, Gingras G, Brisson M. Modelling multi-site transmission of the human papillomavirus and its impact on vaccination effectiveness. Epidemics 2017;21:80-7.

Malagon T, **Lemieux-Mellouki P**, Laprise JF, Brisson M. Bias Due to Correlation Between Times-at-Risk for Infection in Epidemiologic Studies Measuring Biological Interactions Between Sexually Transmitted Infections: A Case Study Using Human Papillomavirus Type Interactions. Am J Epidemiol 2016;184:873-83.

Lemieux-Mellouki P, Drolet M, Brisson J, Franco EL, Boily MC, Baussano I, Brisson M. Assortative mixing as a source of bias in epidemiological studies of sexually transmitted infections: the case of smoking and human papillomavirus. Epidemiol Infect 2016;144:1490-9.

Drolet M, Benard E, Boily MC, Ali H, Baandrup L, Bauer H, Beddows S, Brisson J, Brotherton JM, Cummings T, Donovan B, Fairley CK, Flagg EW, Johnson AM, Kahn JA, Kavanagh K, Kjaer SK, Kliewer EV, **Lemieux-Mellouki P**, Markowitz L, Mboup A, Mesher D, Niccolai L, Oliphant J, Pollock KG, Soldan K, Sonnenberg P, Tabrizi SN, Tanton C, Brisson M. Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis 2015;15:565-80.

Laprise JF, Drolet M, Boily MC, Jit M, Sauvageau C, Franco EL, **Lemieux-Mellouki P**, Malagon T, Brisson M. Comparing the cost-effectiveness of two- and three-dose schedules of human papillomavirus vaccination: a transmission-dynamic modelling study. Vaccine 2014;32:5845-53.

Liste of Presentations

Lemieux-Mellouki P., Drolet M., Brisson M., Impact of changes in sexual behavior on past and future trends of HPV infections and related cancers. Eurogin 2018. Lisbon, Portugal, December 3, 2018 (Oral presentation)

Drolet M., **Lemieux-Mellouki P.**, Mondor M., Fournier A., Markowitz L., Brotherton J., Kreimer A., Brisson M., Reduced-dose HPV vaccination effectiveness in post-vaccination surveillance studies: using mathematical modeling to quantify the impact of biases, 32th International Papillomavirus Conference. Sydney, Australia, October 2018 (Poster presentation)

Lemieux-Mellouki P., Drolet M., Brisson M., Modelling multi-site transmission of HPV and its impact on vaccination effectiveness against HPV? 30th International Papillomavirus Conference. Lisbon, Portugal, September 21, 2015 (Oral presentation)

Lemieux-Mellouki P., Drolet M., Brisson M., Can the association between Human papillomavirus (HPV) and smoking be the result of ignored confounding? 29th International Papillomavirus Conference. Seattle, Washington, August 20-22, 2014 (Poster presentation and discussion)

Lemieux-Mellouki P., Multisites model of transmission vs classical model of genital transmission: implication for the population-level efficacy of the HPV vaccine. CHU de Québec Graduate Student Conference. Québec, May 5, 2014 (Oral presentation)

Lemieux-Mellouki P., Systematic review of oral HPV risk factors. 28th International Papillomavirus Conference. San Juan, Puerto Rico, November 30-31 & December 1-6, 2012 (Poster presentation)

List of Tables

List of Figures

List of Abbrieviations

AIN: Anal intraepithelial neoplasia
ASCUS: Atypical squamous cells of undetermined significance
CIN: Cervical intraepithelial neoplasia
DNA: Deoxyribonucleic acid
HIM: Human papillomavirus infection in men (cohort)
HIV: Human immunodeficiency virus
HPV: Human papillomavirus
HR: High oncogenic risk
HSIL: High-grade intraepithelial lesion
HSV: Herpes simplex virus
LR: Low oncogenic risk
LSIL: Low-grade intraepithelial lesion
MSM: Men who have sex with men
NHANES: National Health and Nutrition Examination Survey
NHSLS: National Health and Social Life Survey
NSFG: National Survey of Family Growth
OR: Odds ratio
SCC: Squamous cell carcinoma
SES: Socioeconomic status
STI: Sexually transmitted infection
US: United States (of America)

Avant-propos

Cette thèse porte sur des questions de recherche qui s'inscrivent dans l'effort actuel de prévention de l'infection au VPH, dont le directeur de thèse, Marc Brisson, est un acteur important sur le plan international. Les questions de recherche ont été réfléchies et formulées selon les besoins concernant l'utilisation de la modélisation mathématique pour l'évaluation de l'efficacité populationnelle de la vaccination contre le VPH. Cette recherche a été financée par les Instituts de recherche en santé du Canada (IRSC) et le CHU de Québec. La thèse comporte trois articles scientifiques dont la description suit.

Le premier article s'intitule : "*Assortative mixing as a source of bias in epidemiological studies of sexually transmitted infections: the case of smoking and human papillomavirus* ". Cet article a été publié dans Epidemiology and Infection le 20 novembre 2015. J'en suis le premier auteur et les autres auteurs sont: Mélanie Drolet, Jacques Brisson, Eduardo Franco, Marie-Claude Boily, Iacopo Baussano et Marc Brisson. M. Brisson et moi avons conçu le devis de l'étude. J'ai codé le modèle mathématique, réalisé les simulations ainsi que les analyses, et rédigé le manuscrit. M. Brisson et M. Drolet ont co-rédigé le manuscrit.

Le deuxième article s'intitule : "*Modelling multi-site transmission of the human papillomavirus and its impact on vaccination effectiveness* ". Cet article a été publié dans Epidemics le 21 décembre 2017. J'en suis le premier auteur et les autres auteurs sont: Mélanie Drolet, Mark Jit, Guillaume Gingras et Marc Brisson. M. Brisson et moi avons conçu le devis de l'étude. J'ai codé le modèle mathématique, réalisé les simulations ainsi que les analyses, et rédigé le manuscrit. M Brisson et M. Drolet ont co-rédigé le manuscrit.

Le troisième article s'intitule : "*Impact of changes in sexual behavior on past and future trends of HPV infections and related cancers* ". Cet article n'a pas encore été soumis pour publication. J'en suis le premier auteur et les autres auteurs sont: Mélanie Drolet et Marc Brisson. M. Brisson et moi avons conçu le devis de l'étude. J'ai codé le modèle mathématique, réalisé les simulations ainsi que les analyses, et rédigé le manuscrit. M Brisson et M. Drolet ont co-rédigé le manuscrit.

Introduction

0.1 Background and rationale

HPV is a sexually transmitted infection which causes cervical cancer[1], a proportion of anogenital[2] and head-and-neck[3] cancers, and anogenital warts[4]. Since 2007, vaccination programs against HPV infection have been implemented in more than 100 countries and territories[5;6]. Cervical cancer is responsible for most of the burden of HPV-associated cancers[7]. For this reason, publicly-funded vaccination programs were initially introduced for girls to reduce incidence of cervical cancer, pre-cancerous lesions, and collaterally anogenital warts.

Transmission-dynamic models, which capture herd immunity effects, have been essential in informing policy-makers on the effectiveness and cost-effectiveness of different HPV vaccination strategies[8]. These models can simulate the transmission process of HPV and predict changes in HPV infections and related diseases after HPV vaccination. These predictions from models are affected by uncertainty regarding our knowledge of the natural history of infection and of the sexual behavior of the targeted population. More specifically, we identify two key issues in current HPV models: 1) HPV is modelled as a uni-site infection, although it is a multi-site infection 2) sexual behavior is assumed to be stable in models, but sexual behavior is known to have changed in the last century.

Firstly, because the focus of prevention has been on genital diseases (e.g., cervical cancer, genital warts), current HPV models only include transmission of genital infections, and do not include transmission between genital and extragenital sites (e.g., oral cavity). However, there is evidence of transmission between these different sites (e.g. oral sex being a mode of transmission between oral and genital sites). By ignoring extragenital transmission of HPV, current HPV models may be producing biased predictions of the long-term impact of HPV vaccination on HPV infections.

Secondly, HPV models do not account for the changes in sexual behavior over the past century, and thus indirectly assume that transmission has reached an equilibrium state. Currently, the incidence of most STIs has not reached a steady state (equilibrium)[9]. This has been attributed, at least in part, to changes in population-level sexual behavior[9]. Precisely, it has been observed that the number of sexual partners has increased and the age of sexual initiation has decreased since the beginning of the

century[10]. It has also been suggested that the sexual revolution of the 1960s resulted in an increase of some sexual practices such as oral and anal sex[11;12]. These changes in sexual behavior are hypothesized to be the main cause of the substantial rise in anal and oropharyngeal cancers in the past two decades[12;13]. For instance, in Canada, oropharyngeal and anal cancers have increased by 20% and 30% respectively, during the span of less than two decades[14;15]. The increase in oropharyngeal cancer affects predominantly white males, in whom the incidence has increased more than 2-fold in the same period[16]. Concerns of a continued "epidemic rise in cancer" are further fueled by extrapolation of current trends predicting that incidence of HPV-positive oropharyngeal cancer will exceed the number of cervical cancers by 2020[16]. However, future trends cannot be anticipated by linear projections from past increases in cancer incidence; projections should be based on the changes in sexual behavior over time, which seems to have stabilized in the past two decades[17;18]. Hence, including the potential effect of changes in sexual behavior in HPV models addresses an ignored source of uncertainty in population-level HPV vaccination effectiveness predictions and will produce projections of future trends in HPV-related cancers that accounts for non-linear changes in sexual behavior over time.

Assessing the changes in sexual behavior is complicated by our poor understanding of many aspects of sexual behavior. In fact, sexual behavior includes not only the acquisition of new partners and the sexual practices, but also "who-mixes-with-whom" or in other words the topology of the sexual network (sexual mixing). For instance, individuals may choose partners similar to them (assortative mixing) with respect to important causal factors of transmission, such as age, level of sexual activity, and even tobacco consumption. Assortative mixing has been shown to be of great importance in the spread of STIs and is commonly included in HPV models. Other factors like smoking and alcohol consumption have received less attention even though they could also play a role in transmission dynamics. Assessing the potential effect of assortative behaviors and changes over time in these behaviors could provide new insight on trends in HPV infection. In particular, assessing the impact of assortativity on HPV transmission could be crucial in understanding the causal roles of risk factors such as smoking in HPV transmission and HPV-related cancers in the past, present and future. Indeed, assortativity according to risk factors such as race is a strong determinant of the distribution of an infection in the population[19]. Consequently, assortativity can be a strong source of confounding in epidemiological studies of HPV risk factors.

0.2 Objectives

The objectives of this project are to develop three different HPV transmission-dynamic models to:
1) Assess the impact of including multi-site transmission to current uni-site HPV models on predictions of the population-level effectiveness of HPV vaccination.
2) Examine how changes in specific aspects of sexual behavior such as mixing, rates of new partner acquisition, sexual practices, age of sexual debut, may have impacted trends in HPV-related oropha-

ryngeal and cervical cancers incidence since 1970, and predict future trends in these cancers.
3) Determine conditions under which assortative mixing could cause a bias in studies examining the causal role of risk factors of STIs and quantify the magnitude of this potential bias.

0.3 Litterature review

0.3.1 Epidemiology of HPV infection

Overview

More than 40 HPV genotypes can infect the mucosa of the anogenital tract, the oral cavity and the respiratory tract[20;21]. These HPV types are classified as HR and LR[20]. Infection with HR-types has been shown to be a necessary cause of cervical cancer[22], but is also a cause of a subset of other anogenital tract and head and neck cancers. In North America, it is estimated that HR-HPV types are responsible for around 50-70% of oropharyngeal cancers[3;7;16], 90% of anal cancers[7;23], 40% of vaginal cancers, vulvar and penile cancers[23]. Among the HR-types, HPV16 causes an estimated 72% of all HPV-associated cancers[23]. Infection by LR-types can cause anogenital warts, mild dysplasia, and recurrent respiratory papillomatosis[4]. More than 90% of anogenital warts are caused by HPV6 and HPV11[4].

HPV infects the stem cells in the basal epithelial layer[24]. Hence, unlike other STIs like HIV , HPV is a localized infection, which makes it important to distinguish between the epidemiology of infection at the different sites.

Table 0.1 provides a summary of some aspects of the epidemiology of HPV infections at the genital, oral and anal sites.

Table 0.1 – **Epidemiology of HPV infection at the genital, oral and anal sites**

Sites	**Prevalence**†		**Seroconversion**		**Duration of Infection**††		**Risk factors**†††	
	Women	Men	Women	Men	Women	Men	Women	Men
Genital	27% [25]	45% [26]	60% by 18mo after clearance [27] (HPV16)	13% by 24mo after clearance [28] (any HPV) 19% by 36mo after clearance [29] (any HPV)	10-11mo [30;31] (HPV16)	7.7-12.2mo [32;33] (HPV16)	Sexual activity: higher number of sexual partners, young age at sexual debut [25;34–36] Smoking [37–39] Contraceptive use [37;39;40] Young adult [41]	Sexual activity: higher number of sexual partners, not using condoms [26;42–44] Smoking [26;42] Alcohol [26;42] Uncircumcised [26;42;43;45;46]
Oral	4% [47]	10% [47]	Unknown	Unknown	3.5mo [48] (HPV16)	7.3-7.5mo [48;49] (HPV16)	Sexual activity: higher number of vaginal sex, oral sex, and deep-kissing partners [47;50;51] Smoking [47;52] HIV [53]	Sexual activity: higher number of vaginal sex, oral sex, and deep-kissing partners [47;50] Smoking [47] HIV [53]
Anal	27-31% [54;55]	16% [56]	Unknown	6% by 36mo after clearance [29] (any HPV)	5mo [57] (HR-HPV)	Unknown	Sexual activity: higher number of sexual partners, anal sex practice [54;55] Smoking [55] Genital HPV [58–60]	Sexual activity: higher number of sexual partners, receptive anal intercourse [56;61;62] HIV [63–65]

† Prevalence in the general population, †† Median, ††† Risk factors of HR-types

Genital infection

Genital HPV infection in women can be detected from exfoliated cells of the vulva, the vagina and the cervix. Infection at the cervix and vagina is common, with an estimated prevalence of 27%[25] among the general female US population (aged 14 to 59 years old). Genital HPV infection among women has been demonstrated to be predominantly sexually transmitted[66]. Acquisition of genital HPV has very rarely been observed among virginal women in longitudinal cohort studies[34;67], suggesting that non-sexual modes of transmission are not significant. In addition, various measures of sexual activity such as a greater number of sexual partners and younger age at sexual debut have been consistently associated with prevalent and incident genital HPV[25;34–36]. Reflecting this, the prevalence of HPV in the cervix peaks at the beginning of adulthood and then decreases with age in HIC[41;68]. The peak of prevalence coincides with the first years of sexual activity, which are characterized by high rate of partner change and susceptibility to infection. A second peak of prevalence has been observed in women older than 45 years old[69]. The steep decrease in prevalence following the initial peak has been interpreted as the result of clearance of infection and protection from reinfection due to acquisition of natural immunity. The immune mechanisms responsible for protection to reinfection are not completely understood, but serum type-specific antibodies have been shown to play an important role in natural immunity[70]. Antibodies are genotype-specific, and an estimated 60% of women seroconvert within 18 months of a detected genital HPV16 infection[27]. Seroconversion has been observed to be protective against genital reinfection of the same type among young women but not among older women[71]. This observation along with the second peak of incidence around 45 years old is compatible with a reactivation of latent infections in older women[72], or with the acquisition of new infections in older women in whom natural immunity waned. The median duration of genital HPV16 infection in women is around 10 and 11 months[30;31], more than 70% of these infections are transient and clear within 24 months of detection[30]. The interpretation of detection and clearance is limited by the inability to differentiate clearance of infections, infections becoming latent and reinfection between study visits[73;74]. It is also unknown whether an individual is infectious for the full duration of infection as viral load has been observed to fluctuate significantly in the course of an infection[31].

Other risk factors associated with prevalent cervical HR-HPV infections include smoking and oral contraceptive use. However, it is not clear whether smoking is associated with increased prevalence independently of sexual activity[37–39], or if it is simply a correlate of sexual behavior or sexual network. The results of studies are inconsistent regarding the association between oral contraceptive use and prevalent cervical HPV infection, with some studies showing no association[39], increased risk with contraceptive use[37], and even decreased risk with contraceptive use[40].

Genital HPV infection in men is commonly detected on the scrotum and the penis (shaft, glans, prepuce, and urethra). In a systematic review of heterogeneous populations, the majority of studies have observed a prevalence of genital HPV among men above 20% (2%-35%)[75]. Prevalence of genital HPV in men was measured at 50% in the HIM study[32]. The ongoing HIM study is conducted among

a cohort of men recruited from the general population of Brazil, Mexico and Tampa Bay. This cohort has been followed for around a decade to study the natural history of HPV in men. In the general US population aged 18-59 years, the penile HPV prevalence was 45%[26]. Similarly to the situation in women, risk of genital HPV in men increases with the number of sexual partners[32;75] and lower age at sexual debut[75]. Smoking and alcohol drinking are also associated with prevalent penile HPV[26;42], but it is unknown if these factors are predictors of prevalent infections independently of sexual activity. The association between penile HPV and circumcision has also frequently been assessed due to the known protective effects of circumcision against other infections. The results are not completely consistent, with some studies showing strong associations[43;45;46], while other studies show no association[26;42]. The association between penile HPV and condom use is puzzling: men reporting intermittent condom use have higher HPV prevalence than men who report never using condom, even after adjustment for lifetime number of sexual partners[26;42]. Furthermore, men who report always using condom do not consistently have lower prevalence than men who report never using condom. Age-specific incidence and prevalence seem to vary less in men compared to women[32;76]. Seroconversion following an incident genital infection is much less common among men than women (13% within 2 years)[28]. Furthermore, antibodies have not been found to be protective against subsequent detection of HPV DNA on the penis[77;78]. These findings could explain the lower prevalence of serum antibodies in men than women[79] and the lesser variation in age-specific HPV prevalence/incidence among men. It is also possible that the detection of HPV DNA on the penis is less often the result of a true active infection. Median duration of genital HPV16 infection was observed to be 7.7 months and 12.2 months in two different studies[32;33].

Oral infection

Prevalence of oral HPV in the general US population (14 to 69 years old) has been estimated at 10%[47] for men and 4%[47] for women, significantly lower than the prevalence at the genital sites of both men and women. The three-fold higher prevalence of oral HPV among men than among women persists even after adjustment for other demographic variables and sexual behavior variables[47]. The leading hypothesis for this difference is that women more often acquire natural immunity from genital infections, which protects them from future oral infections[48]. The overall lower prevalence of oral HPV infections compared to genital HPV infections appears to be mostly the result of lower incidence rather than lower duration of infection[80;81]. Oral HPV has consistently been observed to be more prevalent in smokers[47;52] than in non-smokers, with or without co-variables adjustment (including measures of sexual activity). Smoking has also been observed to be associated with both acquisition of oral HPV infection and longer duration of infection[52;82], but not consistently and the evidence is sparse. Hence, it is still unclear if the higher prevalence in smokers is the result of greater susceptibility to HPV acquisition, longer duration of infection or even confounding by uncontrolled variables related to sexual activity[83]. HIV infection is also associated with higher prevalence and incidence of oral HPV infection[82;84]. The duration of these infections may also be longer for HIV-infected individuals[53]. Another possible risk factor is alcohol drinking[47;49;50;82], but its association with incident or prevalent

oral HPV is inconsistent throughout studies. Tonsillectomy could be protective of oral HPV infections[82], but it has not received much attention in studies. Epidemiological studies have shown that sexual activity is a key risk factor of HPV infection at the oral site in both men and women[47;51;85]. The prevalence of oral HPV is particularly rare among virgin teenagers and adults (0.9%)[47], highlighting the importance of sexual activity in transmission of oral HPV. Oral sex is believed to have an important role in transmitting the infection to the oral cavity, but the determination of specific sexual modes of transmission is complicated by the correlation in sexual practices. Yet, an association between oral HPV risk and the number of oral sex partners has been observed[47;51], even after adjusting for the number of vaginal sex partners[50;51]. Barrier use during oral sex is also associated with less prevalence of oral HPV after adjustment[86]. There is also similar, but weaker, evidence that deep kissing could be a mode of transmission of oral HPV[50;51;85]. In the US, the age-specific prevalence of oral HR-HPV follows a bimodal distribution, the first peak at around 30 years and the second peak at around 60 years[47].

It is not clear at the moment what kind of immune response elicit oral HPV infections, although there are some evidences suggesting that infection of the tonsils could cause seroconversion[79]. Median clearance time of oral HPV infection in the HIM cohort was 6.3 months for HR-types[49], suggesting that most infections are transient, similar to the situation at the genital site (although latent infections are again a possibility).

Anal infection

Infection by HPV at the anal canal has been much less studied, except in some specific populations (HIV+ individuals, MSM). Nevertheless, anal HPV infection has been shown to be common in both men and women. The HIM study[56] measured a prevalence of anal HPV of 16% among men from the general population of Mexico, Brazil and Florida, while two studies measured a prevalence of 31% among young women in Costa Rica[55] and of 27% among adult women in Hawaii[54]. Variables measuring the level of sexual activity such as the number of sexual partners have been associated with anal HPV in these studies in both men and women. The practice of anal intercourse has also been associated with anal HPV prevalence among women[54;55], and thus could be an important mode of transmission to the anus. Smoking was observed to be a risk factor of prevalent anal HPV among women independently of measures of sexual activity[55]. In women, having a genital HPV infection is also a risk factor for prevalent anal HPV infection[58–60], and this could be due to autoinoculation. In MSM, receptive anal intercourse is a risk factor of anal HPV[61;62]. Furthermore, MSM have significantly higher prevalence of anal HPV than Men-who-have-sex-with-women (MSW)[56]. HIV-positive MSM have also a higher prevalence than HIV-negative MSM[63;64;84] even after adjustment for measures of sexual activity. The age-specific prevalence of anal HPV was observed as being relatively stable for both women and men[54–56].

In a study of MSM[87], anal HPV was observed to be the main predictor of seropositivity to type-concordant HPV antibodies, while penile HPV showed no association with seropositivity and oral

HPV showed a non-significant association with seropositivity. Although the results of this study are not applicable to the general population, one explanation suggested by the authors is that infection at mucosal sites has a higher chance of inducing seroconversion than infection at keratinized epithelium like the penile shaft. In the HIM cohort, rates of seroconversion were 7% within 36 months of anal HPV clearance. In the cohort of adult women from Hawaii[57], median clearance of anal HPV was 5 months for HR-types.

Transmission and interaction between sites

HPV is thought to access the basal layer of the epithelium through micro abrasions[24]. Hence, direct contact with the infected mucosa could transmit the infection if such micro lesions exist. As previously mentioned, there are strong evidences of an important role for penetrative sex (i.e., anal, vaginal and oral sex) as modes of transmission between sites[35;36;47;51;88]. However, HPV could be transmitted from one site to another through various other intimate contacts or even non-sexual contacts[89]. In a longitudinal study of heterosexual couples, correlations between infections at sites not involved in sexual intercourse has frequently been observed[88]. Hand-genital transmission between partners and scrotum-penile self-inoculation could have occurred. Self-inoculation is suggested to be the result of hand-genital contacts. Hands could thus be a bridge for infection between two sites. However, detection of HPV DNA in this study could be the result of contamination, rather than true infection. Another study found that HPV detection on the genitals was highly correlated with detection on the hands, but HPV presence on the hands was more likely the results of self-inoculation rather than hand-genital or hand-hand transmission between partners[90]. Other studies have reported that cervical HPV infection caused a higher risk of incident and prevalent type-concordant anal HPV infection[58–60], even in the absence of anal sex. Non-penetrative modes of transmission such as autoinoculation from vaginal shedding or intimate contacts during foreplay were suggested[58]. Toilet wiping behaviors[60] were also found to be associated with anal HPV[60] in women with previous HPV-related genital neoplasia. In a study of women, self-inoculation behaviors (e.g., putting fingers in mouth after having touched genitals) were found to be predictive of oral HPV independently of measures of sexual activity[91]. A meta-analysis of studies assessing HPV co-infection at the oral and cervical sites has concluded in a greater-than-expected prevalence of these co-infections, supporting dependence between oral and genital infections[92]. However, such dependence should not be over-interpreted, as it could be the result of simply having an infected partner.

Challenges and biases in the study of HPV infection epidemiology

Assessing the causal role of risk factors of HPV infections Factors can be causally related to prevalent HPV through various mechanisms: 1) increased susceptibility to acquiring new infections upon contact with an infected (e.g., HIV), 2) increased duration of HPV infections, 3) increased exposure to HPV infected individuals (e.g., high number of sexual partners, sexual activity of partners). Investigators are often interested in identifying factors "biologically"-related to HPV, which typically means factors that increase susceptibility or duration of HPV infections. Prevention of these biolog-

ical causes will have a straightforward and predictable beneficial effect (e.g., HIV antiviral treatment to reduce HPV duration) for the individual and indirectly the whole population. On the other hand, causes related to HPV exposure and sexual network may have complicated and unpredictable effects. Hence, to identify the biological causes, it is necessary to adjust for exposure to HPV infection. In studies, this is often done by adjusting for measures of sexual activity at the individual-level, such as the participants' lifetime number of sexual partners. Yet, because HPV is a STI, the risk factors of the sexual partners also need to be considered to accurately adjust for the risk of exposure to HPV. This is sometime referred as the "male-factor"[93–95] in studies of risk factors of cervical cancer and cervical HPV infections, but this is rarely investigated due to the difficulties involved in acquiring information on sexual partners. Therefore, in many analyses of risk factors of prevalent HPV, such as smoking, confounding by risk factors of sexual partners (e.g., sexual activity) is a possibility that is not considered. The potential for confounding by exposure to HPV has been an important issue in the assessment of the inter-dependence between sites of infection and HPV types. In several studies, having an infection at one site (e.g., genital site) was found to be a risk factor for having an infection at another site[35;59;92]. Similarly, having an infection of one type (e.g., HPV11) or having another STI, is often found to be a risk factor of having (or acquiring) an infection of another type (e.g., HPV16)[96–98]. However, these data cannot be directly taken as evidence, for example, of synergy between HPV-types or between other STIs and HPV. In fact, it may be that having a STI is associated with having sexual partners at high-risk of any STI. Furthermore, it has been shown that correlation in the presence of infections of different HPV-types or sites of infection is to be expected simply due to the fact that all infections are transmitted through the same mode: sexual contacts[99]. In summary, obtaining conclusive results from analyses of HPV risk factors is a task complicated by the lack of adjustment for HPV exposure.

Understanding the natural history of HPV There are still many uncertainties and unknowns regarding the aspects of the natural history of HPV infection relevant to transmission: the duration of infectiousness, the variation in infectiousness during the course of an infection, the possibility for the virus to become latent and reactivate, the possibility of being re-infected during the course of an infection, the possibility of acquiring local and systemic immunity to reinfection. Originally, the prevailing paradigm was that most HPV infections cleared after about a year, infected individuals were contagious during that period, and could develop natural systemic immunity (serum antibodies) upon clearance[72]. However, results of longitudinal studies on the natural history of HPV infection are difficult to interpret. Because HPV is mostly asymptomatic, identifying acquisition and clearance of infection is based on HPV-DNA detection in exfoliated cells taken from the site of infection. However, a positive HPV-DNA result may not always reflect an active infection, but rather a deposit of HPV-DNA. In fact, HPV can even contaminate environmental surfaces[100]. Therefore, what is counted as clearance of infection in studies may, in some cases, be instead clearance of a deposit. The infectiousness potential of these deposits is also unknown[89]. Furthermore, clearance of infection is determined from a negative HPV-DNA test, but shedding of HPV-infected cells may diminish or even cease with-

out true clearance of infection. Variation in shedding of HPV-infected vaginal cells was observed to be correlated with the menstrual cycle[101]. Hence, dynamic immune-related factors could explain a negative HPV-DNA test without actual clearance of infection. The hypothesis of infections frequently becoming latent and reactivating has gained credibility[72].

Longitudinal studies of HPV acquisition in individuals testing positive or negative to HPV serum antibodies have not provided epidemiologists with a clear understanding of natural immunity to HPV. Naturally acquired serum antibodies have not consistently been found to be protective against acquisition of HPV[72]. It is possible that detection of a new HPV infection in a seropositive woman is the result of the reactivation of a latent infection which may also have caused her seroconversion. It is also unclear if naturally acquired antibodies protect systemically against infections at all sites, and if there exists site-specific immunity.

Finally, little is understood about HPV infectiousness other than it could be associated with detected HPV viral load. Viral load of HPV detected on the penis was found to be associated with concordance of the infection in the female partner[102]. It is not known if viral load remains sufficiently high for HPV to be transmitted throughout the whole course of an infection. It is plausible that viral load, and thus infectiousness, will vary throughout the course of infection according to dynamic immune-related factors.

0.3.2 Changes in sexual behavior pertaining to HPV transmission

Many aspects of sexual behavior have changed during the last century. Because HPV is predominantly sexually transmitted, these changes in sexual behavior have likely affected the epidemiology of HPV across birth cohorts.

Trends in vaginal sex (sexual intercourse)

An important trend in the US is the decline in age at first vaginal sex among women, which can be traced back to birth cohorts from the 1910s[103]. For instance, around 4% of women born in the 1940s have had sexual intercourse before the age of 15, whereas the proportion of teenage girls who have had sexual intercourse in the 14-15 age-group was 13% in 2009[104]. Another marked trend is the increase in the number of heterosexual partners, which can be observed between cohorts born before 1930 up to cohort born in 1970[103;105]. In 1990, the proportion of women born before the 1930s who have had one or less lifetime number of sexual partners was 62% compared to 38% for women born in the 1950s, even though these younger women had less time to accumulate partners. Between cohorts born in 1940-49 and cohorts born in 1980-89, median lifetime number of partners has increased from 2.6 to 5.3 in women and from 6.7 to 8.8 in men[10]. In the same period, age at first sexual intercourse decreased from 17.9 to 16.2 years in women and from 17.1 to 16.1 years in men[10]. These changes were observed to be correlated with the reorganization of the course of sexual life, in particular the lengthening of the premarital period[103]. In recent years, studies conducted among populations of

teenagers have provided evidence that the trend of decreasing age at sexual debut in the past decades may be stabilizing, or even reversing[106–109]. In particular, the proportion of teenagers who ever had intercourse declined during the 90s, and was stable during the 2000s[106;108].

Trends in oral and anal sex

The practice of both oral and anal sex has only recently been receiving attention in national surveys. Yet, changes in these sexual practices are important as they may have affected the rates of transmission of STIs. Two non-representative surveys conducted in 1940 and 1967 in populations of young college-educated individuals have highlighted an increase in the practice of oral sex in the premarital period[11]. In a study from France, an increase in the practice of oral and anal sex was observed between 1970 and 2006[12]. In women aged 60-69 years, around 40% of women had experienced cunnilingus in 1970 compared to around 70% in 2006. The proportions were similar for men experiencing fellatio[12]. These trends have also been partially observed in the US[110]. In a study comparing data from the NSFG in 2002 with two studies conducted in 1991, an increase in the proportion of individuals practicing oral and anal sex through the 90s as well as in the number of sexual partners was observed[110]. While oral sex is now a common practice, regular practice of anal sex is still marginal[110]. In the US, 89% of women aged 25-29 have performed fellatio and 46% have had anal sex in their lifetime, yet 50% have performed fellatio in the past month[104] and only 5% have had anal sex in the past month. Although there have been concerns of a recent increase in the practice of oral sex among teenagers[111], current evidence does not support a strong change in the practice of oral sex among teenagers. The National Survey of Adolescent Males has provided nationally representative data on this phenomenon between 1988 and 1995[112]. In this study, the proportion of virgin males who ever had oral sex did not increase significantly, except among blacks teenagers[112]. A comparison of the 2002 with the 2006-2010 NSFG showed no increase in the practice of oral sex among teenagers[105].

Biases in trends in sexual behavior

The interpretation of data on trends in sexual behavior is complicated by at least two potential biases: 1) reporting bias in sexual behavior surveys (e.g., due to social/sexual norms[113]), 2) survivor bias in older individuals if promiscuous behavior is associated with greater mortality. Both these biases would be expected to result in lower reported sexual activity in older birth cohorts of women. In fact, social norms and attitudes toward sexuality have shifted from conservative to liberal throughout the century[18], and therefore it was less accepted for older birth cohorts of women to be promiscuous during their youth. Furthermore, sexual activity of older birth cohorts has to be assessed through the reporting of very old participants, which may result in a survivor bias if participants alive are less likely to be promiscuous. In men, these biases may be in the opposite direction, since there is a possible double standard in how sexual norms apply to men and women[114].

These potential differences between men and women in the nature and the effect of sexual norms may also explain the notable discrepancy in reported sexual partners among men and women[115], with

men reporting systematically greater number of female partners. In the UK population aged more than 16 years, the mean number of lifetime partners is 14.14 in men and 7.12 in women, a 2-fold difference[116]. The discrepancy is greater in older birth cohorts[115], which would be coherent with stricter cultural pressure and sexual norms, and with a resulting greater reporting bias in these older birth cohorts[115]. Hence, one possible explanation for the discrepancy is that men tend to overestimate their number of sexual partners and women tend to underestimate their number of sexual partners. Supporting this, researchers have observed that the average number of sexual partners reported by women was 2.6 when they thought someone could know their answer and 4.4 when they thought a lie could be detected by the investigators[117]. In men in the same situations, the average number of sexual partners went from 3.7 to 4.0. In men, overestimation may be the result of the higher prestige associated with having many sexual partners[118], or greater frequency of rounding up the number of sexual partners due to a more skewed distribution[116]. It may also stem from the under-sampling and under-reporting of highly sexually active females such as sex workers[119]. In fact, sex workers could have hundreds, if not thousands of sexual partners in their lifetime[119]. Hence, even a low percentage of sex workers could shift the mean number of sexual partners considerably[119]. Yet, in many surveys, the lifetime number of sexual partners is right-censored at a much lower value (e.g., 50 in the NSFG). The results of a recent analysis[116] suggested that the discrepancy is due to men rounding up, men reporting extreme values, and differences in sexual attitudes, rather than the under-sampling of sex workers. However, the estimation of the contribution of sex workers to the discrepancy was based on men's reports of paid partnerships in a sexual behavior survey. Yet, men may under-report paid encounters since these could decrease their prestige.

Epidemiology of HPV-associated cancers and prevention

Cervical cancers Cervical cancer accounts for most of the burden of HPV-associated cancers with 12 000 new cases per year in North America[7]. Cervical HPV infection is a necessary cause of cervical cancers[7]. Genotypes 16 and 18 are involved in >70% of the cases[22]. Although most (>70%) cervical HPV infections are cleared naturally within 2 years, a minority of these infections will persist, and among the persistent infections few will progress to invasive cancer. The period of progression from infection to cancer can last from 10 to 30 years, during which pre-cancerous lesions will form and progress in grade of dysplasia[120]. Grades of lesions are divided as mild cervical dysplasia (cervical intraepithelial neoplasia, CIN1), moderate cervical dysplasia (CIN2), and severe cervical dysplasia (CIN3). Most CIN1 lesions will clear without intervention: the proportion regressing within 2 years of detection is around 50%[121]. The CIN2/3 lesions have higher chance of persisting and leading to cancer, but around 35% of these lesions will regress to CIN1 or less within 2 years of detection[121]. Several characteristics of the host and of the virus are co-factors in the progression to cancer. Smoking, taking oral contraceptives, and high number of childbirths have all been implicated in cancer progression[122–124]. Non-attendance to cervical cancer screening is also one of the key risk factor[125]; this is expected as screening was the only effective prevention tool before the introduction of vaccination. The risk of progression to cancer is also highest for HPV16[126]. Risk of cancer for a specific

genotype could vary by geographic region due to intratypic genetic variation[127].

Head-and-neck cancers In North America, there is roughly the same number of new oropharyngeal cancer cases in men as of new cervical cancer cases per year ($\approx$ 12 000 cases)[7]. It is estimated that at least 60% of these cancers involve HPV and >85% of HPV-associated oropharyngeal cancers involve HPV16[3]. HPV is an established causal factor of the SCC of the base of the tongue and tonsils[128;129], which account for over 90% of oropharyngeal cancers[130]. Other oropharyngeal cancers and oral cavity cancers do not involve HPV as frequently[131;132], and the role of HPV in these cancers is still controversial. For oral cavity cancers, the etiologic fraction was estimated at 6%[132]. The HPV-negative head-and-neck cancers have different etiological profiles, involving notably smoking and alcohol consumption, diet and poor oral hygiene[133]. Evidences are not consistent as to whether smoking and alcohol also play a role in HPV-associated oropharyngeal cancer[133;134]; smoking and alcohol could be factors of progression from infection to cancer or factors of oral HPV acquisition[47;133;135]. Tonsillectomy is also associated with a 85% reduction in tonsil cancer incidence among individuals younger than 60 years old[136]. Because the practice of tonsillectomy has been under scrutiny in recent years[136], there was a possibility that a decrease in the practice of tonsillectomy could explain trends in oropharyngeal cancer. However, a recent analysis has shown that it is unlikely that temporal trends in tonsillectomy had a strong effect on temporal trends of tonsils cancer[137]. A higher number of oral sex partners is strongly associated with incident HPV-positive oropharyngeal cancer[133;138], and this association is likely the result of the association between oral sex and oral HPV acquisition. HPV-positive cancer patients are 3 years younger in average than patients with HPV-negative cancer[14]. The incidence of HPV-associated oropharyngeal cancer is 4.9 times higher in men than women[139], and this could be explained by the higher prevalence of oral HPV prevalence in men (10%) than in women (4%). In the US, the incidence of oropharyngeal cancer has increased by approximately 20% since 1980[140], but the increase is more marked among white males: more than 2-fold since 1980. The increase of HPV-positive oropharyngeal cancers is partially masked by the decline in HPV-negative oropharyngeal cancers. The actual increase in HPV-positive oropharyngeal cancers was estimated to be more than 3-fold[16].

Anal cancers Around 4000 new cases of anal SCC are diagnosed per year in North America[7;141]. Infection by HPV is a causal factor of SCC of the anus[141;142], and is detected in 90% of these cancers[7]. In particular, HPV16 is detected in around 65% of invasive anal cancer[143]. Similar to cervical cancer, anal SCC has precursor lesions (AIN), which are graded according to the severity of dysplasia (AIN1/2/3)[144]. Data on progression and regression rates of lesions are currently lacking, but the high prevalence of anal HPV coupled with the low incidence of anal cancer suggests that fewer anal HPV infections progress to cancer compared to cervical HPV infection[144]. Anal cancer disproportionately affects certain sub-groups of the populations. Women who have had cervical cancer are at increasing risk of anal cancer[141], which could be due to the role of the cervix as a reservoir for HPV infection at the anal canal[58;60]. MSM with and without HIV have respectively a 31-fold and 16-fold increased incidence of anal SCC[145]. The association between MSM and anal cancer could be partly mediated

by the association between receptive anal intercourse and anal cancer[146]. Smoking has also been identified as an important risk factor of anal cancer[146]. In the US, the incidence of anal cancer has increased by approximately 60% since 1975[147].

A summary of the epidemiology of HPV-associated cancers is given in Table 0.2.

Table 0.2 – **Burden and risk factors of HPV-associated cancers**

	Proportion due to HPV in the US	**Incidence in the US (per 100 000)**[148]	**Risk factors**
Cervical cancer	All HPV: 100%[7] HPV16: ≈ 50%[149]	7.4	Sexual activity, smoking, oral contraceptive, parity, screening attendance[122;123;125]
Oropharyngeal cancer	All HPV: ≈ 60%[7] HPV16: >55%[3]	4.3 in men 0.9 in women	HPV-related: Sexual activity[133] HPV-unrelated: smoking, alcohol, diet[133]
Anal cancer	All HPV: ≈ 90%[7] HPV16: ≈ 60%[143]	1.6 in men 2.3 in women	HPV related: Sexual activity, immunosuppression (e.g., HIV), smoking[145;146]

0.3.3 Prevention of HPV-related cancers and diseases

Screening

Screening for cervical cancer was introduced in the 1950s with the Pap test. Until recently, precancerous cervical lesions were detected using cytology (e.g., Pap test or liquid cytology), and then lesions were confirmed by colposcopy. The severity of the lesion (LSIL, ASCUS, HSIL) required to be referred for colposcopy varies across screening programs and physicians. HPV-DNA testing has replaced cytology as the first-line test in some countries (e.g., Australia[150], Argentina[151]), with cytology being used in case of positive HPV-DNA results to determine whether a woman should be retested later or should undergo colposcopy. HPV testing has also been introduced for triage of ASCUS and LSIL after cytology (e.g., Denmark[152], Norway[153]). HPV testing works by detecting existing HPV infections with high-risk types (e.g., HPV16) in the cervix which are necessary for the occurrence of lesions. This method has proven more sensitive than cytology, albeit at the cost of a loss of specificity[154]; this is why triage with cytology is useful in case of positive HPV testing. Confirmed lesions can be treated by loop electrosurgical excision procedure, cryotherapy or cold knife conization[155]. The burden of cervical cancer has been greatly reduced in past decades due to the success of screening and treatment of precancerous lesions. In the US, the incidence of cervical cancer has decreased by 54% between 1973 and 2007[156].

At this time, there is no screening for potential oral or anal precursor lesions, unlike for cervical cancers. The equivalent of the Pap test for oropharyngeal cancer yielded inconclusive results in a study[157]: no association between the presence of HPV16 and cytological abnormality was observed.

Because anal cancer is rare, screening would likely not be cost-effective in the general population. Even among high-risk populations (e.g. HIV+ MSM), the cost-effectiveness of screening is unclear, because of uncertainty around the natural history of anal cancer and around the effectiveness of the various possible treatments of AIN lesions[158].

Vaccination

Three prophylactic vaccines are currently available: the bivalent, quadrivalent, and nonavalent vaccines which protect against respectively HPV16 and -18, HPV16, -18, -6, -11, and HPV16, -18, -31, -33, -45, -52, -58, -6, -11. All three vaccines have demonstrated high efficacy (>98%) in preventing cervical HPV infections and subsequent precursor lesions associated with HPV types included in the vaccines[159–161]. The duration of protection cannot be determined yet, but no evidence of loss of protection has been observed[162], suggesting long duration of protection. For the bivalent vaccine, the duration of protection exceeds 10 years[163]. The vaccines have also been shown to have efficacy against oral and anogenital infections in men[164;165].

Around 40% of the countries have introduced vaccination programmes against HPV, with more than

80% of the HIC having done so[5;166]. The quadrivalent vaccine is the most commonly used in the programmes (50% of the market), but the newly introduced nonavalent vaccine has replaced the quadrivalent vaccine in many HIC (28% of the market)[5]. Preadolescent girls (<12 years old) are the primary target for vaccination with schedules varying from 2-doses to 3-doses[166]. The World Health Organization also recommends vaccination of multiple age cohorts of girls aged between 9 and 14 years old for faster and broader protection at the start of the program[167]. Because of the financial cost of mass immunization and the lower burden of male HPV-diseases, vaccination has been extended to boys in some countries, but not all: the US, Switzerland, Australia Canada, Israel, Austria are all vaccinating boys. Importantly, the list of countries vaccinating boys is growing, and the decisions of extending vaccination to boys are being made based on considerations of gender-equity in addition to effectiveness.

0.3.4 Models of HPV transmission and effectiveness of vaccination

Mathematical models integrate data on sexual behavior of the population and the natural history of HPV infection and diseases to produce simulations of HPV transmission with and without vaccination. Such simulations are currently the main method to obtain estimates of population-level effectiveness and cost-effectiveness of vaccination. In fact, estimates from randomized trials (randomized at the individual-level) of vaccine efficacy cannot account for herd immunity[168]. Herd immunity is the indirect protection non-vaccinated individuals receive when transmission is reduced by the vaccination of a large enough proportion of the population. Without accounting for herd immunity, estimates of efficacy and population-level effectiveness of vaccination can be underestimated.

More than 50 studies have been published on the effectiveness and cost-effectiveness of vaccination against HPV infection[169]. In HIC, these analyses have consistently shown that vaccinating adolescent girls is cost-effective regardless of the type of vaccine[8;170;171]. Vaccination of women older than 18 is not predicted to be cost-effective partly because the proportion of women infected by HPV prior to vaccination would be too high[171]. Vaccination of boys may be cost-effective in a context of low coverage among girls and low cost of vaccine dose[8;170;171]. For the same price, the nonavalent vaccine has been shown to be more cost-effective than the quadrivalent vaccine even when assuming a shorter duration of protection and slightly lower efficacy[172]. Two-dose vaccine schedule is predicted to be more cost-effective than three-dose schedule if the duration of the protection from two doses is at least 20 years[173;174]. Likewise, one-dose schedule is predicted to be more cost-effective than two-dose schedule if the duration of protection from one dose of vaccine is more than 20 years[175].

Static vs dynamic models

To perform the analyses of effectiveness and cost-effectiveness, static and dynamic models have been used[8]. Static models do not integrate the dynamic aspect of transmission: the incidence rate of infection among the susceptibles is assumed constant and independent of the number of infected. Thus, static models cannot account for herd immunity, the indirect protection unvaccinated individuals gain

through the vaccination of others. For this reason, static models will produce conservative estimates of the benefits of vaccinating girls, which can be sufficient for policy-makers. However, transmission-dynamic models are necessary to assess the benefits of adding boys to the vaccine programme[8], as herd immunity effect from girls vaccination may reduce the incremental gain of adding boys. Likewise, transmission-dynamic models are necessary to compare the effectiveness of different vaccine schedules or strategies, such as the reduction in the number of doses or catch-up vaccination of older females.

Transmission dynamic models of HPV infection

Transmission-dynamic models are mechanistic models of transmission: they include sexual partnership formation, transmission within partnerships and the subsequent course of infection among the infected. These models can be individual-based or compartmental depending on the degree of resolution needed. Individual-based models are more convenient when a high degree of heterogeneity is required for the model agents.

The transmission-dynamic models of HPV developed to inform current vaccination policy decisions have followed the standard *SIS*[176] (susceptible $\rightarrow$ infected $\rightarrow$ susceptible), *SIR*[176;177] (susceptible $\rightarrow$ infected $\rightarrow$ recovered/resistant) or *SIRS*[178;179] paradigm. The different HPV genotypes are ideally modelled individually, because grouping multiple HPV-types together as one infection will incorrectly reproduce herd immunity[180]. The most common HPV-types modelled are 16, 18, 6 and 11, which are responsible for the majority of the burden and are targeted by the quadrivalent vaccine. Models have also included HPV-types targeted by the nonavalent vaccine: 31/33/45/52/58[172;181].

A core element of these transmission-dynamic models of HPV infection is the simulation of the sexual activity of the model population, which will cause the virus to spread. To reproduce sexual activity, the population is typically stratified according to the level of sexual activity and the age of individuals[182]. Individuals in the model form sexual partnerships at different rates depending on their level of sexual activity and age. With whom individuals form partnerships is specified through a matrix of probabilities called "Who-mixes-with-whom"[183]. A common feature of such matrix is assortativity: the greater likelihood that two individuals forming a partnership share similar characteristics (e.g., similar age). HPV transmission models also have to account for the demographic characteristics of the simulated population (e.g., the distribution of age). In a recent meta-analysis on predictions from transmission-dynamic HPV models, the sources of heterogeneity in the predictions of different models were identified[169]. Among other factors, stratification of the population by level of sexual activity was identified as having a great impact on predictions.

Calibration of HPV transmission-dynamic models

To simulate HPV transmission, models depend on the parameters governing sexual behavior (e.g., sexual partner acquisition rate, who-mixes-with-whom matrix), and on the parameters of natural history

of infection and diseases (e.g., duration of infection). These parameters are informed by calibrating the models to a multitude of data sources [184] including sexual behavior surveys, cross-sectional studies of HPV prevalence or seroprevalence, and longitudinal studies of HPV incidence and clearance. These empirical data have been used to produce calibration targets such as the number of sexual partners per year, age-specific prevalence, seroprevalence or incidence of HPV16, of CIN lesions and of cervical cancer [184]. Several statistically rigorous methods for calibrating models to data have been employed: likelihood maximization [185], Bayesian computation [178] and Approximate Bayesian Computation [186]. These methods can all account for parameter uncertainty which stems from variance in simulations (if model is stochastic) and in the calibration data. Uncertainty around the parameters of the vaccine programme (e.g. vaccine coverage) or parameters uninformed by data (e.g., parameter of assortativity) has been assessed through sensitivity analyses [186].

Specificities of HPV transmission-dynamic models

Closing the population HPV models, unlike many other infectious diseases models, are not typically calibrated to time-series of case count data. This is because most HPV infections are asymptomatic and require DNA sampling for detection. Hence, it is logistically difficult to get incidence data on a large sample. Rather, infection frequency in the general population is measured as prevalence of infection or prevalence of antibodies, as in NHANES. The calibration to prevalence data only rather than incidence data is more complex. Indeed, prevalent infection or seroprevalence can be the consequence of sexual activity that occurred long ago which means that models should technically simulate sexual behavior and transmission that occurred decades before occurrence of prevalence data and sexual behavior data. Furthermore, HPV has been present in human populations for thousands of years, so a recent epidemic period with a baseline null HPV prevalence cannot be identified unlike for infections like a new strain of flu or HIV. In summary, with HPV models, there is the issue of closing the population with respect to transmission: defining a population for which we have data and such that transmission does not depend on individuals outside the population (e.g., individuals of previous generations).

To circumvent this issue, an assumption of stability is frequently implicit in HPV models [186;187]. It is assumed that the observed HPV prevalence, HPV transmission, and the sexual behavior of the population is at equilibrium; the sexual behavior of the current older cohorts is thus the future sexual behavior of the current younger cohorts. With these assumptions, a baseline HPV prevalence is obtained by simply simulating transmission with the model until an equilibrium state is reached.

HPV is a localized infection Because cervical cancers account for an estimated 87% of all HPV-attributable cancers worldwide [7], the main focus of HPV related research and prevention has traditionally been cervical cancer and anogenital warts. Hence, most models used to answer questions related to the effectiveness and cost-effectiveness of different HPV vaccination strategies and schedules have included cervical cancer and anogenital warts as possible diseased states following an HPV infec-

tion[182;187], but few have included oropharyngeal or anal cancers. Furthermore, to our knowledge, none of the models of heterosexual transmission of HPV used for policy decisions developed over the past 15 years have incorporated extragenital infections and multi-site transmission, although models including MSM can include at least two sites (the anal or penile site)[?]. Indeed, previous models of heterosexual transmission were "uni-site" models, where infection is only acquired and transmitted at one site in women (cervix/vagina) and men (penis)[188]. Few of these uni-site models have also been used to predict the reduction of non-cervical HPV-related cancers[173;189] (e.g., oropharyngeal cancers) due to vaccination. In these analyses, all HPV-related cancers are assumed to stem from the same uni-site infection in both men and women. Hence, the transmission of infections at the extragenital sites that results in extragenital cancers is not reproduced, which may cause bias in predictions of reduction of these cancers. Furthermore, the dynamics of transmission at the cervicovaginal site may be different in the presence of multi-site transmission[190]. As a consequence, estimates of vaccination effectiveness against cervical cancer could change with the inclusion of multi-site transmission. Importantly, the inclusion of multi-site infections in transmission models comes with additional questions and complexities regarding natural history of infection. For instance, it becomes crucial to distinguish between site-specific and systemic natural immunity.

Interestingly, the problem can also be posed for multi-site infections within a specific macro-site[191]. For instance, cervical HPV infections may be better represented as infections of specific patches within the cervix, rather than the whole cervix. This may also have implications for transmission dynamics as an infected individual could get infected at another patch, and infectiousness could depend on the size of the area infected. In brief, the localized nature of HPV infections is not captured by the current HPV models of natural history.

Using transmission-dynamic models to understand biases in epidemiological studies As mentioned in 0.3.1, results of studies on the epidemiology of HPV can be difficult to interpret. This is in part because HPV is an infectious disease, and the methodology used in studies is adapted to non-transmittable outcomes (outcome for which the stable-unit-treatment-value assumption (SUTVA) is plausible[192]). Transmission-dynamic models can be used as the complex causal system that generates the data observed in HPV studies. Doing this replaces the violated SUTVA with the assumptions underlying the transmission-dynamic model used. Within this new system, the data observed in the studies may be reanalyzed, and potential biases in the original analysis can be assessed. For instance, this methodology has been used to show that the data used as evidences against cross-immune interactions between HPV-types may be reproduced in a transmission-dynamic model even in the presence of cross-immune interactions[99]. Simulations from transmission-dynamic models can also be used to help determine optimal study design with respect to statistical power and magnitude of bias[193–195].

0.3.5 Summary of methods

We developed different models to address our three objectives. The general characteristics of these models are presented in Table 0.3. Importantly, our models were not used to produce accurate predictions of effectiveness or cost-effectiveness of interventions, but rather to understand three specific phenomena.

Table 0.3 – **Description of the models used in the thesis**

	Models type	**Model stratification**	**Modelled outcome**	**Calibration targets**	**Outcome measures**
Objective 1	• Dynamic • Deterministic • Compartmental • SIR^a	• Gender • Level of sexual activity	• Genital HPV16 • Extragenital HPV16^b	HPV16 prevalence (Genital, oral and anal)	Prevalence ratio of genital HPV16 post-vaccination compared with pre-vaccination
Objective 2	• Dynamic • Stochastic • Individual-based • SIR^a	• Gender • Age • Year of birth • Sexual life history	• Genital HPV16 • Oral HPV16 • HPV16-related oropharyngeal cancer • HPV16-related cervical cancer	• HPV16 prevalence (Genital, oral) • HPV16-related cancers incidence (cervical and oropharyngeal)	Incidence of HPV16-related oropharyngeal and cervical cancers
Objective 3	• Dynamic • Deterministic • Compartmental • SIR^a	• Smoking status • Level of sexual activity	• Genital HPV16	HPV16 prevalence (Genital)	Odds ratio of HPV prevalence between smokers and non-smokers

^{a} : Susceptible→Infected→Recovered natural history, ^{b} : multi-site model only

The first objective of this thesis is to assess the impact of including multi-site transmission to current uni-site HPV models on predictions of the population-level effectiveness of HPV vaccination. Hence, we developed for this objective a model of multi-site HPV transmission in addition to a uni-site transmission model. Both models were based on ordinary differential equations. The models formulation can be found in the Appendix A and are described in Chapter 1. The uni-site model simulated HPV transmission only at the genital site (between the penis and the cervico-vaginal site), while the multi-site model simulated transmission at the genital site and at one extragenital site representing either the oral or the anal site. Furthermore, HPV transmission between the genital and extragenital sites was included. Thus, there were four possible pathways of transmission in the multi-site model: genital→extragenital, genital→genital, extragenital→extragenital, extragenital→genital. HPV infections were cleared according to a rate specific to each site. Upon clearance of infection, there was a probability of developing natural immunity. We considered different scenarios regarding the effect of natural immunity: site-specific immunity provided lifelong protection only for one site, and systemic natural immunity provided lifelong protection for both sites. We estimated the probabilities of transmission per partnership by fitting the models to age-specific HPV16 prevalence at the genital and extragenital sites (i.e., oral and anal). Finally, we predicted the population-level effectiveness of girls-only HPV vaccination by estimating the percent reduction in genital HPV16 prevalence at post-vaccination equilibrium (i.e., more than 100 years after the start of vaccination).

The second objective is to examine how changes in specific aspects of sexual behavior such as mixing, rates of new partner acquisition, sexual practices, age at sexual debut, may have impacted trends in HPV-related oropharyngeal and cervical cancer incidence since the 1970s, and predict future trends in these cancers. We developed for this objective a stochastic individual-based transmission-dynamic model. Detailed description of the model can be found in the Appendix B and C, and is described in Chapter 2. Briefly, the model population included individuals born between 1850 and 1999 and aged between 12 and 44 years old. We reproduced their sexual history, which included: sexual partnerships occurring in the context of cohabitation, sexual partnerships occurring outside cohabitation, and whether oral sex was practiced within each partnership. Sexual activity was determined based on data from the NHSLS, NSFG and GSS and three surveys from France. Then, we simulated HPV16 transmission at the genital and oral sites from 1862 (the year when those born in 1850 turn 12) to 2015. Transmission was based on two probabilities: 1) the probability of acquiring a genital infection within a partnership with someone infected at the genital site, 2) the probability of acquiring an oral infection within a partnership with someone infected at the genital site and on whom oral sex is practiced. HPV infections cleared according to a rate and, upon clearance, there was a site-specific probability of developing lifelong systemic immunity. Infected individuals could progress to either oropharyngeal or cervical cancer, according to two parameters that were gender- and site-specific: 1) the probability of a newly acquired infection to progress toward cancer, 2) the dwell time between the acquisition of infection and the diagnosis of cancer for infections that progress toward cancer. We estimated the two probabilities of HPV16 transmission by fitting HPV16 prevalence in 2015 at the oral site in men and

at the genital site in women among the US population aged between 20 and 30 years. We estimated the parameters of progression from infection to HPV16-related cervical and oropharyngeal cancers by fitting the age-specific incidence of these cancers in the US. We used the years 1973-1975 for cervical cancer incidence and the years 1984-1985 for oropharyngeal cancer incidence. With the calibrated model, we produced predictions of the incidence of HPV16-related oropharyngeal and cervical cancers (without screening) between 1915 and 2045. Figure 0.1 illustrates the process and the data used.

The third objective of the thesis is to determine conditions under which assortative mixing could cause bias in studies examining the causal role of risk factors of STIs and quantify the magnitude of this potential bias. We developed a deterministic HPV16 transmission-dynamic model using ordinary differential equations. Only one site of infection is included in the model. The model formulation can be found in the Appendix D and is described in Chapter 3. The model included stratification for two levels of sexual activity and two levels of smoking habits (smokers and non-smokers). There were two important characteristics of the model population: 1) smokers had higher level of sexual activity, 2) sexual mixing was assortative with respect to smoking status. The latter means that smokers had higher chance of forming partnerships with smokers (and symmetrically for non-smokers). Model parameters included the probability of HPV16 transmission per partnership, HPV16 clearance rate, and the probability of developing natural immunity upon clearance, the degrees of assortativity by smoking status and level of sexual activity. All parameters, but the degrees of assortativity, were based on prior modelling work. The degrees of assortativity were assumed in the main analysis but were varied extensively in sensitivity analyses due to their uncertainty. We then simulated the conduct of a cross-sectional study in the model population in which the association between smoking and HPV16 prevalence is estimated as an odds ratio of HPV16 prevalence. The measure of association is estimated both with and without adjustment for sexual activity of the participants. In this fictive study, sexual activity is perfectly adjusted for and there is no other potential confounder at the individual-level. Furthermore, smoking is not a biological cause of infection in the model, which means that the only possible reason for a deviation from the null association would be a bias due to assortativity with respect to smoking status. We termed this bias *assortativity bias*.

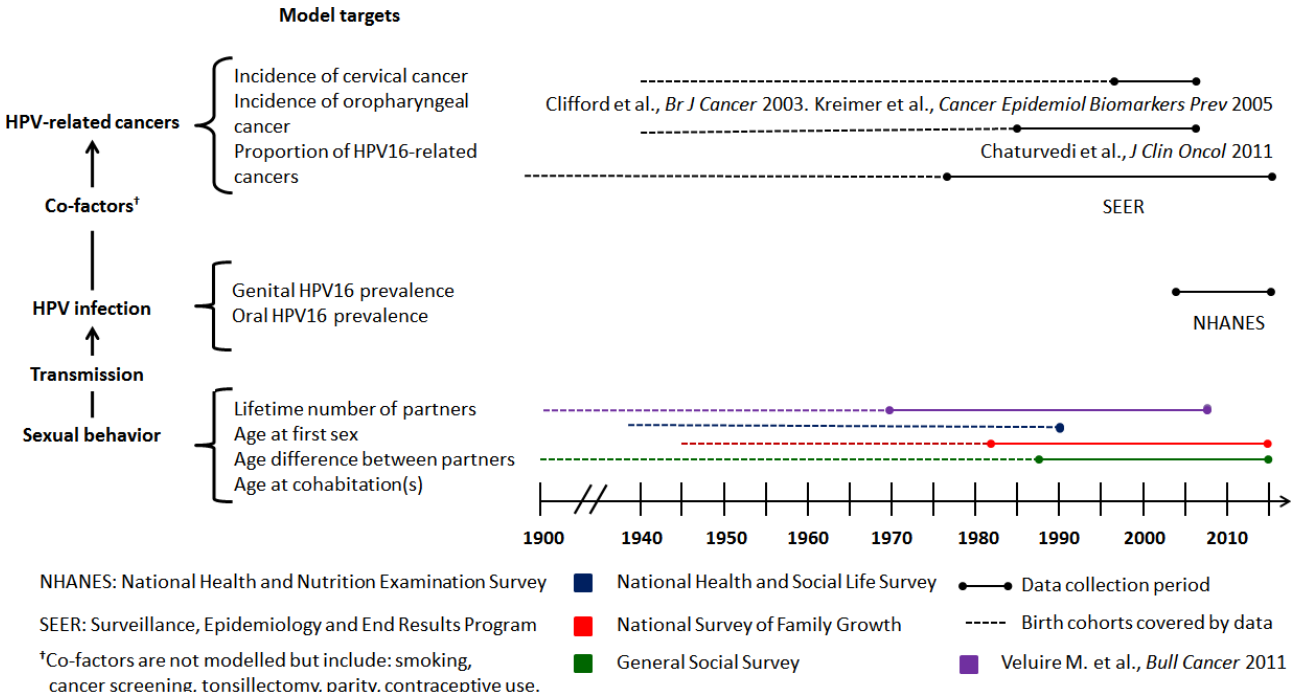


Figure 0.1 – **Summary of the methodology and data used for objective 2.** A mathematical model was developed which reproduces sexual behavior, HPV transmission and development of cancer in the last century. The model was calibrated to empirical data on sexual behavior, HPV infection prevalence, and HPV-related cancers incidence.

Chapter 1

Modelling multi-site transmission of the human papillomavirus and its impact on vaccination effectiveness

1.1 Résumé

Traditionnellement, les modèles de transmission des virus du papillome humain (VPH) incluaient uniquement la transmission génitale. Nous avons comparé les prédictions d'efficacité populationnelle de la vaccination contre l'infection génitale au VPH16 chez les femmes, avec un modèle uni-site (génital), et un modèle multi-site (génital et extra-génital). Nous avons calibré ces modèles à la prévalence de VPH16 (génital: 5%-7.5%; extra-génital: 0%-7.5%). L'immunité naturelle pouvait être systémique (tous les sites) ou spécifique à un site. L'issue était la réduction relative de la prévalence de VPH16 génital chez les femmes post-vaccination (RR_{prev}). En supposant une immunité spécifique, RR_{prev} était généralement supérieur avec le modèle multi-site que l'uni-site. Cette différence était plus importante si la transmission extra-génital→génital était élevée. En supposant une immunité systémique, RR_{prev} était généralement inférieur avec le modèle multi-site. Ces résultats suggèrent que l'efficacité populationnelle de la vaccination prédite avec un modèle uni-site a peu de chance d'être significativement biaisées.

1.2 Abstract

Objective: Previous HPV models have only included genital transmission, when evidence suggests that transmission between several anatomical sites occurs. We compared model predictions of population-level HPV vaccination effectiveness against genital HPV16 infection in women, using a 1) uni-site (genital site), and a 2) multi-site model (genital and one extragenital site).

Methods: We developed a uni-site and a multi-site deterministic HPV transmission model, assuming

natural immunity was either site-specific or systemic. Both models were calibrated to genital HPV16 prevalence (5%-7.5%), whilst the multi-site model was calibrated to HPV16 prevalence representative of oral (0%-1%) and anal (1%-7.5%) sites. For each model, we identified 2,500 parameter sets that fit endemic genital and extragenital prevalences within pre-specified target ranges. In the Base-case analysis, vaccination was girls-only with 40% coverage. Vaccine efficacy was 100% for all sites with lifetime protection. The outcome was the relative reduction in genital HPV16 prevalence among women at post-vaccination equilibrium (RRprev). RRprev was stratified by extragenital prevalence pre-vaccination.

Results: Under assumptions of site-specific immunity, RRprev with the multi-site model was generally greater than with the uni-site model. Differences between the uni-site and multi-site models were greater when transmission from the extragenital site to the genital site was high. Under assumptions of systemic immunity, the multi-site and uni-site models yielded similar RRprev in the scenario without immunity after extragenital infection. In the scenario with systemic immunity after extragenital infection, the multi-site model yielded lower predictions of RRprev than the uni-site model.

Conclusion: Modelling genital-site only transmission may overestimate vaccination impact if extragenital infections contribute to systemic natural immunity or underestimate vaccination impact if a high proportion of genital infections originate from extragenital infections. Under current understanding of heterosexual HPV transmission and immunity, a substantial bias from using uni-site models in predicting vaccination effectiveness against genital HPV infection is unlikely to occur.

1.3 Introduction

Human papillomavirus (HPV) is a sexually transmitted infection (STI), able to infect the basal epithelial layer of the cervix, oral cavity, the anus and the genitals. The main focus of HPV related research and prevention has historically been cervical cancer, for which HPV is a necessary cause. This is mainly because cervical cancers account for an estimated 87% of all HPV-attributable cancers worldwide[7]. However, research on non-cervical HPV infections and disease has dramatically increased since 2005. Two main reasons explain this intensified focus on non-cervical HPV: 1) a steep increase in the incidence of oropharyngeal and anal cancers in the US and other high income countries[7;196] and 2) recent results showing that HPV vaccines are highly effective at preventing persistent HPV infection and pre-cancerous lesions in sites other than the cervix[197–200].

Despite the recent focus on non-cervical HPV research, there remain significant gaps in knowledge, particularly around HPV transmission to and immunity between cervical and non-cervical sites. The few epidemiological studies on multi-site HPV infection/transmission suggest that autoinoculation within one host, or inter-site transmission between individuals may occur[88;201]. Plausible modes of inter-site transmission include oral sex, anal sex, or indirect transmission through contact with hands.

Autoinoculation between the genital and oral or anal sites could occur through intermediate contact with the hands[60;91] or through virus shedding in the anogenital region[58]. Therefore, HPV infection at one site is likely dependent on transmission from other sites. As for natural immunity, studies suggest that production of antibodies is much more frequent following cervical infections than non-cervical infections[27;29]. However, it is unclear whether antibody response is synonymous with systemic protection against subsequent infections at other sites. Furthermore, the role of local immunity, either humoral or cell-mediated, in protecting against subsequent infections is not well understood. Hence, there could be site-specific differences in immune response and vulnerability to subsequent infections.

None of the 19 HPV transmission-dynamic models developed over the past 10 years to assess HPV vaccination effectiveness[202] have incorporated multi-site infections/transmission, which may have biased their predictions. Indeed, all previous models were "uni-site" models, where infection is only acquired and transmitted at one site in women (implicitly the cervico-vaginal region) and men (implicitly the penis). Furthermore, the bulk of previous models were only fit to age-specific HPV infection data at the cervico-vaginal site[188]. By ignoring other potential markers of infection and sources of transmission from extragenital infections, these uni-site models may be biased in their predictions of long term post HPV vaccination dynamics (e.g., herd effects and population-level effectiveness).

Given that the predictions of previous HPV models, based on a uni-site transmission paradigm, were highly influential in HPV vaccination policy decisions worldwide[203], it is important to assess the robustness of the predictions to assumptions about multi-site transmission and natural immunity. The objectives of this study are to: 1) compare predictions of HPV16 vaccination effectiveness and herd effects between multi-site and uni-site transmission-dynamic models, under various assumptions of HPV16 transmission and natural immunity, and 2) understand the effect of the key factors of transmission responsible for difference in predictions of HPV16 vaccination effectiveness between multi-site and uni-site models.

1.4 Material and methods

We developed two multi-site models and one uni-site model to address our objectives.

1.4.1 Comparing predictions of HPV16 vaccination effectiveness between multi-site and uni-site transmission-dynamic models

Model structure

To address objective 1, predictions of HPV16 vaccination effectiveness are compared between a uni-site and a multi-site model. We developed a uni-site and a multi-site deterministic HPV16 transmission model based on the Susceptible-Infectious-Recovered paradigm (see the Supplementary material for the flow diagrams and the model equations). For both models, the population is 1) heterosexual, 2)

open and stable (deaths balance births), and 3) stratified according to gender and two levels of sexual activity. Mixing between levels of sexual activity was assumed to be random. For simplicity, we did not stratify the models by age. On average, individuals spend 15 years in the modelled population, representing the peak years of sexual activity (15-30 years).

The only structural differences between the uni-site and multi-site models are in HPV16 transmission and natural immunity. The uni-site model represents transmission between the cervico-vaginal site and penis, and the probability of natural immunity following clearance is allowed to vary between 0-100% in both women and men. On the other hand, the multi-site model represents the following four transmission pathways: 1) extragenital $\rightarrow$ extragenital, 2) extragenital $\rightarrow$ genital, 3) genital $\rightarrow$ genital and 4) genital $\rightarrow$ extragenital. In the multi-site model, the extragenital site can either be the oral or anal site. Each pathway has its own probability of transmission, which is modeled per sexual partnership (i.e., we did not model duration of sexual partnerships, the specific number of different acts within a partnership or use transmission probabilities per act).

Scenarios with and without autoinoculation between the two sites were investigated. With autoinoculation, individuals infected at one site can get infected at the other site without sexual exposure, according to two time-homogeneous rates corresponding to the two possibilities (genital $\rightarrow$ extragenital and extragenital $\rightarrow$ genital). Given uncertainty in the literature about natural immunity and the possible impact of natural immunity assumptions on predictions, we modelled 4 scenarios. In scenario 1, individuals can only acquire immunity upon clearing genital infection and immunity protects against subsequent genital infections, but not against extragenital infections (*Local immunity after genital infection only*). In scenario 2, individuals can acquire local immunity upon clearing genital and extragenital infections (*Local immunity after genital and extragenital infections*). In scenario 3, individuals can only acquire immunity upon clearing genital infection and immunity protects against subsequent infection at any site (*Systemic immunity after genital infection only*). Finally, in scenario 4, individuals can acquire systemic immunity upon clearing genital or extragenital infection (*Systemic immunity after genital and extragenital infection*).

Parameterization and fitting procedure

To compare vaccination effectiveness predictions between the uni-site and multi-site models, the models were calibrated to the same pre-vaccination HPV16 prevalence at the cervico-vaginal site (prevalence=5.0-7.5%). The lower and upper bounds of HPV16 prevalence were based on estimates from two studies among US women between 14 and 30 years old (around 5.0%[204] and 7.5%[41]). In addition, the multi-site model was calibrated to HPV16 prevalence representing either the oral (prevalence=0.0-1.0%[47;52]) or the anal site (prevalence=1.0-7.5%[56;205]) (see Table 1.1). We chose wide ranges for HPV16 prevalence at the extragenital sites to enable greater generalizability of results. The models were calibrated to HPV16 prevalence by varying HPV16 transmission probabilities from females to males and from males to females. A maximum relative difference of $\pm 15\%$ was

allowed between male-to-female and female-to-male probabilities of transmission. In scenarios with autoinoculation, the two rates of autoinoculation (genital → extragenital and extragenital → genital) were also varied and assumed to be the same for males and females. All other parameters were also identical between males and females and were fixed based on available data in the literature[206] and prior modelling work[207] (see Table 1.1). To select the parameters that produced the best fit to the HPV16 prevalence data, we used a 4 step procedure: 1) each parameter was given a uniform prior (probability of transmission between 0-100%), 2) parameter sets were drawn from the prior distributions using Latin Hypercube Sampling[186;208], 3) parameter sets were selected if they produced HPV16 prevalence estimates within the prespecified target intervals (see Table 1.1), and 4) the calibration procedure was stopped once about 2,500 parameter sets were selected. The uni-site model was calibrated a single time while the multi-site model was calibrated eight times for each of the four different scenarios of natural immunity and the two scenarios of autoinoculation (with or without).

Analysis design and outcome

To investigate the effect of multi-site transmission on estimates of vaccination effectiveness, we modelled a girls-only vaccination scenario, assuming 100% vaccine efficacy against infection (at all modelled sites) and lifelong duration of protection.

For comparisons between the uni-site and multi-site model predictions of vaccination effectiveness, we used the relative reduction in genital HPV16 prevalence at the post-vaccination equilibrium compared to no vaccination. Results are presented using the median, the minimum and maximum, the 25^{th} and 75^{th} percentiles of simulation results using the 2,500 parameter sets identified through calibration.

We assumed vaccination coverage was 40% in the base case, but varied coverage between 0% and 100% in sensitivity analyses.

1.4.2 Understanding the effect of the key factors of transmission responsible for difference in predictions of HPV16 vaccination effectiveness between multi-site and uni-site models.

To address objective 2, we proceeded in two steps:

Step 1: Analysis with a simplified multi-site model

First, a simplified homogeneous multi-site model was used for general tractability and theoretical insights. We identified three key factors responsible for differences in HPV vaccination effectiveness predictions between the multi-site and uni-site models (see Supplementary materials A): 1) the proportion of all incident genital infections that are due to extragenital → genital transmission at pre-vaccination equilibrium (Factor 1: proportion of genital infections caused by extragenital infections);

Table 1.1 – **Uni-site and multi-site model calibration**

	Multi-site model	**Uni-site model**
Calibration target: HPV16 prevalence (females)	Genital [5%-7.5%][41;204] Extragenital [0%-7.5%][47;52;56;205]	Genital [5%-7.5%][41;204]
Scenarios of natural immunity	• Local immunity after genital infection, • Local immunity after genital and extragenital infection, • Systemic immunity after genital infection, • Systemic immunity after genital and extragenital infection	• Immunity after genital infection
Varying parameters	Probabilities of transmission[a]: • Genital → Genital, • Genital → Extragenital, • Extragenital → Extragenital, • Extragenital → Genital Rates of autoinoculation: • Genital → Extragenital, • Extragenital → Extragenital,	Probabilities of transmission[a]: • Genital → Genital
Fixed parameters	Average duration of infection: • 1.5 years (based on cervical HPV[206]) Effective average rate of new partner acquisition per year: • Low level of activity (95%): 1.4 • High level of activity (5%): 5.7 Probability of developing natural immunity after infection[207]: 45%	Same values of fixed parameters as in the multi-site model

[a] Male-to-female and female-to-male probabilities of transmission were allowed to be different (maximum relative difference allowed= $\pm 15\%$). All other parameters were equal between men and women.

this proportion is obtained by dividing the incidence of genital infections caused by the transmission of an extragenital infection to the genital site by the total incidence of genital infections, 2) proportion of extragenital infections caused by genital infections at pre-vaccination equilibrium (Factor 2), 3) proportion of susceptibles to extragenital infections at pre-vaccination equilibrium (Factor 3).

Model structure. The simplified multi-site model follows the same Susceptible-Infected-Recovered structure as the model described in section 1.4.1 (see Supplementary material A for the model equations) with the four transmission pathways modelled as probabilities per instantaneous partnership. However, in contrast, the model includes one level of sexual activity, one gender, no autoinoculation, and transmission from individuals infected at the genital and extragenital sites occur independently (i.e., two independent modes of transmission).

Parameterization and fitting procedure. We used the same values of duration of infection and probability of natural immunity as for the model developed for objective 1 (see Table 1.1). For simplicity, natural immunity was assumed to be local after genital infections (which corresponds to scenario 1 in objective 1).
We aimed to assess the effect of genital $\rightarrow$ extragenital and extragenital $\rightarrow$ genital transmission probabilities on predicted vaccination effectiveness. To do this, we calibrated the four transmission probabilities to targets of 7% for endemic genital prevalence and 3% for endemic extragenital prevalence. These targets were based on HPV16 prevalence targets for objective 1. The four transmission probabilities were calibrated by solving algebraically the model equations to obtain 10 000 parameter sets.

Analysis design and outcome. For objective 2, we used the minimum vaccination coverage needed to eliminate the infection in the population as our main outcome (the elimination threshold, q_c). We estimated the elimination threshold from the basic reproductive number (R_0). For the simple multi-site model, the elimination threshold is given by $\left(1 - \frac{1}{R_0}\right)$. We computed R_0 as the leading eigenvalue of the Next-Generation-Matrix[209].

Step 2: Analysis with the heterogeneous multi-site model

In step 2, we assessed the effect of Factors 1, 2 and 3 on predicted HPV16 vaccination effectiveness using the heterogeneous multi-site model described in section 1.4.1. To achieve this, we calculated, before vaccination, from all the parameter sets identified during the calibration process in objective 1: the proportion of genital infections caused by extragenital infections (Factor 1), the proportion of extragenital infections caused by genital infections (Factor 2), and the proportion of susceptibles to extragenital infections (Factor 3). We then examined the relationships between these outcomes and HPV16 vaccination effectiveness.

1.5 Results

1.5.1 Comparing predictions of HPV16 vaccination effectiveness between multi-site and uni-site transmission-dynamic models

Effect of multi-site transmission on vaccination impact assuming local immunity only after genital infection (scenario 1) or local immunity after genital infection and extragenital infection (scenario 2)

Under the assumption of local immunity after genital infection, the impact of vaccination on the population-level prevalence of genital HPV16 infection predicted by the multi-site model is similar to the uni-site model when extragenital prevalence is low, but the multi-site model predicts substantially greater vaccination effectiveness when extragenital prevalence is high (Figure 1.1A and Table 1.2). The difference is even greater when comparing the 75^{th} quantiles or the maximum predicted effectiveness (Figure 1.1A and Table 1.2). Finally, the inclusion of autoinoculation caused a decrease in predicted effectiveness by around 6 percentage points, assuming an extragenital prevalence between 3.0% and 7.5%.

Under the assumption of local immunity after genital and extragenital infections, predicted effectiveness with the multi-site model is slightly greater than with the uni-site model when extragenital prevalence is low, and slightly lower when extragenital prevalence is high (Figure 1.1B and Table 1.2). As in scenario 1, the distribution of predicted effectiveness with the multi-site model is much more skewed toward higher values (see Figure 1.1B and Table 1.2). Overall, the difference in predictions between the two models is lower in scenario 2 than in scenario 1. Autoinoculation had little impact on predicted effectiveness.

Effect of multi-site transmission, assuming systemic immunity after genital infection only (scenario 3) or systemic immunity after genital and extragenital infection (scenario 4)

Under the assumption of systemic immunity after genital infection, predictions with both models are almost identical with or without autoinoculation (Figure 1.1C and Table 1.2). Under the assumption of systemic immunity after genital and extragenital infections, predicted effectiveness with the multi-site model is lower than with the uni-site model (Figure 1.1D and Table 1.2). The difference between the uni-site and multi-site models increases with higher extragenital prevalence. Autoinoculation caused an increase in predicted effectiveness up to 5 percentage points, assuming an extragenital prevalence of 3.0%-7.5%.

Table 1.2 – **Predicted effectiveness of vaccination against genital HPV16 infection for the multi-site and uni-site models by scenario and extragenital prevalence**

	Multi-site model				Uni-site model
	Extragenital prevalence				
	[0%-1%[	[1%-3%[	[3%-5%[	[5%-7.5%]	
Local immunity after genital infections					
with Autoinoculation	61%(58-64)	65%(62-71)	67%(63-72)	66%(61-71)	59%(57-62)
without Autoinoculation	61%(58-65)	67%(62-76)	73%(66-81)	72%(66-81)	
Local immunity after genital and extragenital infection					
with Autoinoculation	62%(58-65)	64%(60-68)	61%(59-63)	56%(54-58)	59%(57-62)
without Autoinoculation	62%(58-65)	65%(61-70)	63%(60-66)	57%(59-61)	
Systemic immunity after genital infection					
with Autoinoculation	59%(57-62)	61%(59-64)	60%(57-62)	58%(56-61)	59%(57-62)
without Autoinoculation	61%(58-64)	62%(59-65)	61%(59-64)	60%(58-62)	
Systemic immunity after genital and extragenital infection					
with Autoinoculation	58%(56-62)	57%(54-60)	54%(52-56)	50%(49-52)	59%(57-62)
without Autoinoculation	58%(55-61)	54%(52-56)	50%(49-51)	47%(47-48)	

In the base-case, vaccination program is girls-only with 40% coverage and vaccine is assumed to have 100% efficacy and lifelong duration for the two sites of infection. Models predictions are presented as the median and intervals denote the 25^{th}-75^{th} percentiles of predictions.

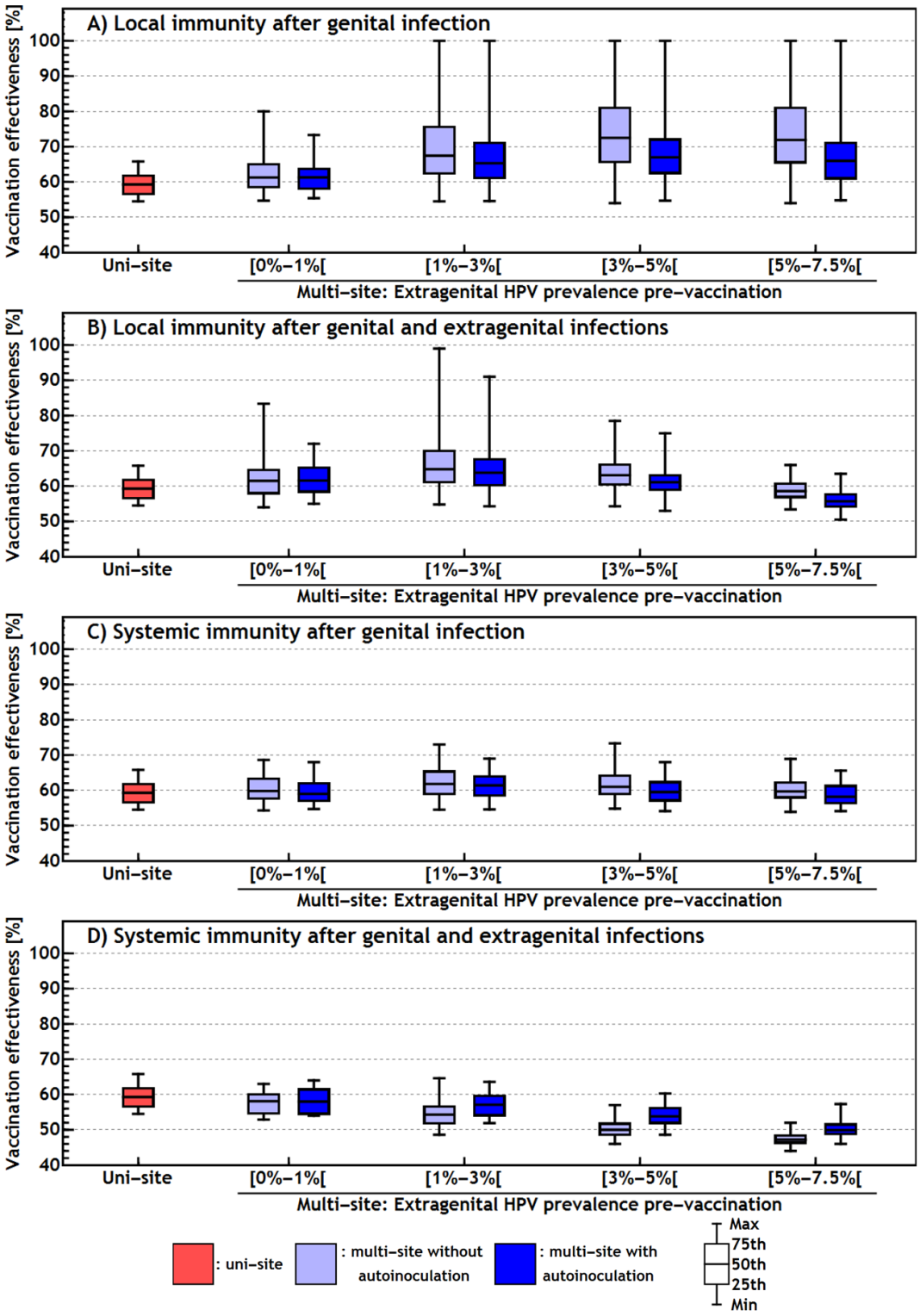


Figure 1.1 – **Population-level vaccination effectiveness of HPV16 vaccination with a multi-site and uni-site models: comparison by prevalence of extragenital HPV16.** Vaccination program is girls-only with 40% coverage and vaccine is assumed to have 100% efficacy and lifelong duration for the two sites of infection. Vaccination effectiveness = relative reduction in genital HPV16 prevalence at the post-vaccination equilibrium compared to no vaccination.

Sensitivity analyses

The qualitative differences between the uni-site and multi-site model predictions are not affected by vaccination coverage as long as coverage is below the elimination threshold (see Figure A.6 in Supplementary materials). For example, under the assumption of local immunity after genital infection (scenario 1), differences between the uni-site and multi-site models start diminishing as coverage exceeds 50% (when the upper range of the multi-site model's predictions reach the elimination threshold) and disappear if coverage exceeds 75% (elimination of HPV16 with both the uni-site and multi-site models).

1.5.2 Understanding the effect of the key factors of transmission responsible for difference in predictions of HPV16 vaccination effectiveness between multi-site and uni-site models.

Effect of inter-site transmission on the elimination threshold with the simple homogeneous multi-site model

Figure 1.2 shows that the elimination threshold decreases with increasing extragenital $\rightarrow$ genital transmission (Factor 1) and/or decreasing genital $\rightarrow$ extragenital transmission (Factor 2). It can also be extrapolated from Figure 1.2 that the minimum elimination threshold decreases as the proportion of susceptibles to extragenital infection increases (Factor 3). This result is stated in full and demonstrated in section A.2.3 of the Supplementary materials. Briefly, the minimum value of the elimination threshold for the multi-site model (3% in Figure 1.2) is equal to 1-proportion of susceptibles to extragenital infections. Thus, if the proportion of susceptibles to extragenital infections increases, 1-proportion of susceptibles to extragenital infections decreases and so does the minimum elimination threshold. The maximum value of the elimination threshold (38% in Figure 1.2) is equal to 1-proportion of susceptibles to genital infections. Hence, the maximal value of the elimination threshold for the multi-site model corresponds to the elimination threshold of a uni-site model of the genital site, and the minimal value to the elimination threshold of a uni-site model of the extragenital site. In particular, if the proportions of susceptibles to genital and extragenital infections are the same, the elimination threshold of the multi-site model will be the same as the elimination threshold of the uni-site model.

In the Supplementary materials, we show analytically the results presented above and that they are not dependent on specific parameter values or assumptions of natural immunity.

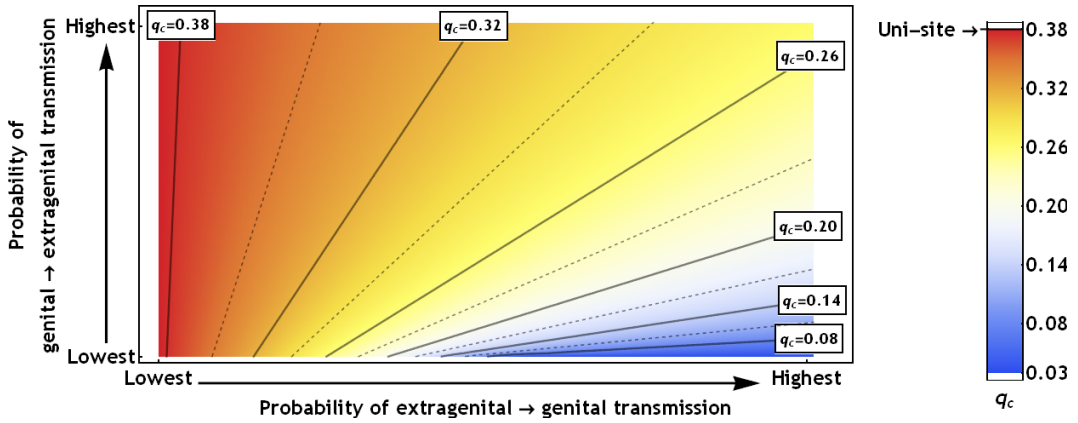


Figure 1.2 – **Effect of transmission parameters on the elimination threshold (q_c) of the multi-site model**. The model was calibrated to a genital prevalence of 7% and an extragenital prevalence of 3%. The x-axis represents probability that a male with an extragenital infection infects the genital site of his partner during a partnership, and the y-axis represents the probability that a male with a genital infection infects the extragenital site of his partner during a partnership. q_c = elimination threshold, minimum vaccination coverage needed to achieve elimination assuming a vaccine with 100% efficacy and lifelong duration. Elimination threshold for the uni-site model is given by the black line on the scale for the elimination threshold.

Effect of inter-site transmission and proportion of susceptibles to extragenital infection on predictions of HPV16 vaccination effectiveness.

Figure 1.3 shows HPV16 vaccination effectiveness predictions of the heterogeneous multi-site model as a function of the three key factors, measured at pre-vaccination equilibrium: 1) the proportion of genital infections that were caused by an extragenital infection, 2) the proportion of extragenital infections that were caused by a genital infection, and 3) the proportion of susceptibles to extragenital infections. The relationships are all monotonic with predicted vaccination effectiveness increasing when the proportion of susceptibles to extragenital infections increases, when the proportion of genital infections caused by extragenital infections increases, and when the proportion of extragenital infections caused by genital infections decreases. These results were the same when including autoinoculation or not.

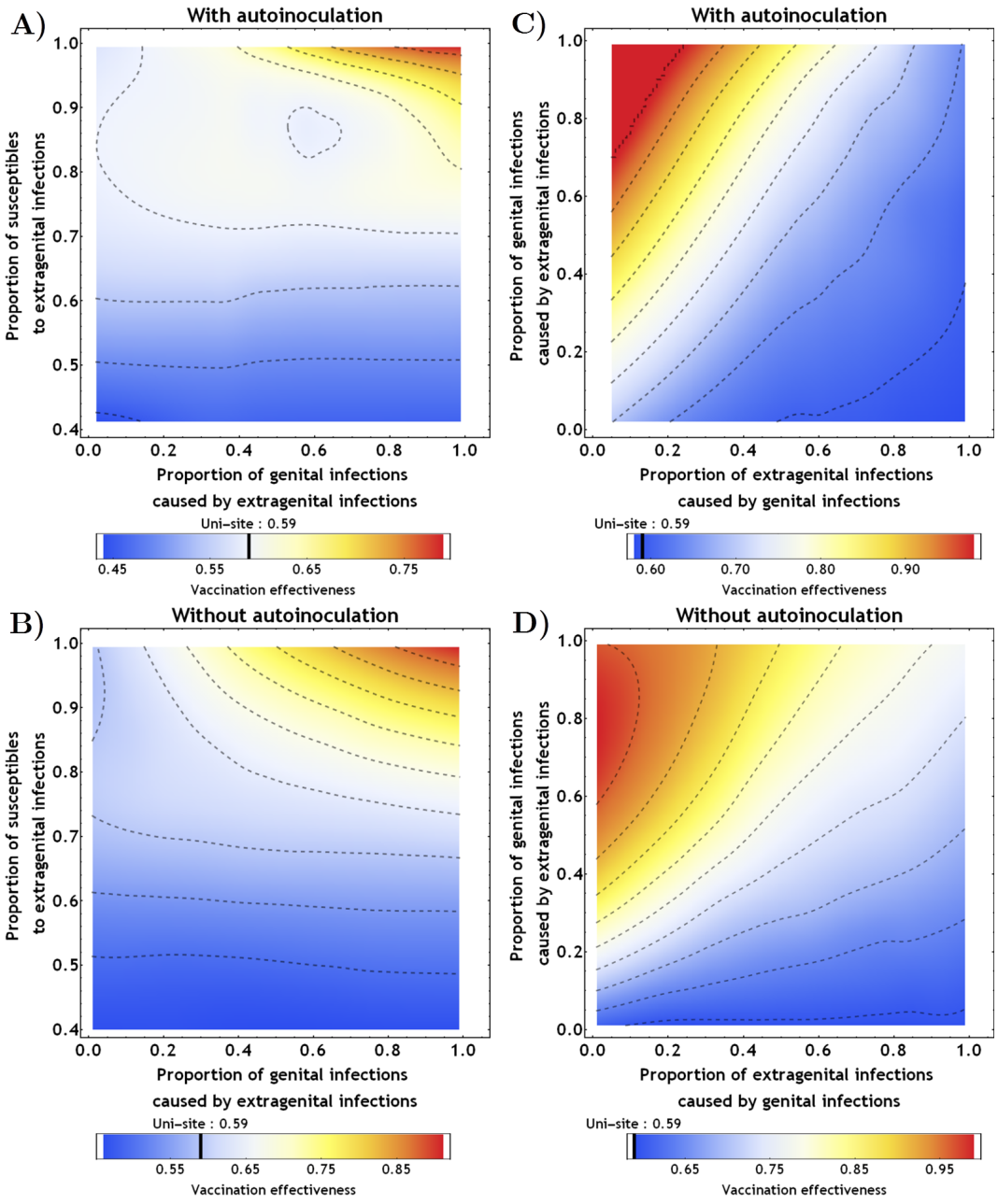


Figure 1.3 – **Effect of inter-site transmission and proportion of susceptibles to extragenital infection on average predicted effectiveness with the multi-site model.** A) Vaccination effectiveness as function of the proportion of susceptibles to extragenital infections and of the proportion of genital infections caused by extragenital infections (with autoinoculation), B) (without autoinoculation), C) Vaccination effectiveness as function of the proportion of genital infections caused by extragenital infections and of the proportions of extragenital infections caused by genital infections (with autoinoculation), D) (without autoinoculation).

IMPORTANTLY: In C) and D) we show the results for simulations where the proportion of susceptibles to extragenital infections is higher than 90% (below this value there was no variability in vaccination effectiveness). Vaccination effectiveness = relative reduction in genital HPV16 prevalence at the post-vaccination equilibrium compared to no vaccination. Proportion of genital infections caused by extragenital infections = (Incidence of genital infections caused by extragenital infections) / (Total incidence of genital infections). Incidence = (contact rate) x (probability of transmission) x (prevalence of infected) x (prevalence of susceptibles to genital infections). OF NOTE: The median prediction of vaccination effectiveness from the uni-site model is given by the black line on the vaccination effectiveness scales. The relation between variables was smoothed through local polynomial regression.

1.6 Discussion

In this paper, we examined whether the predictions of traditional uni-site models that were used to inform decisions about vaccination are biased because they do not take into account transmission between different sites. Our results suggest that the difference between the predictions of the uni-site and multi-site models are a function of natural immunity assumptions and prevalence at the extragenital site. Under the assumption of local immunity (scenario 1 and 2), vaccination effectiveness predictions with the multi-site model are either equal or greater than with the uni-site model. This difference increases when assuming that a greater proportion of HPV16 genital infections was produced by extragenital infections. Under the assumption that natural immunity confers systemic protection against infection at all sites (scenario 3 and 4), the multi-site model predictions of vaccination effectiveness were either the same or lower than the uni-site model predictions.

The effects of natural immunity assumptions are essentially due to differences in the proportions of susceptibles to genital infections and to extragenital infections (Factor 3). The proportion of susceptibles to extragenital infection is the highest under scenario 1 of local immunity after genital infection, because there is no natural immunity to extragenital infections. Predicted effectiveness is consequently highest under scenario 1. The proportion of susceptibles to extragenital infection is lower in scenario 3 (systemic immunity after genital infection) than scenario 2 (local immunity after genital and extragenital infections) and is the lowest in scenario 4 (systemic immunity after genital and extragenital infections). In scenario 3, the proportion of susceptibles to extragenital infection is roughly the same as the proportion of susceptibles to genital infection, which explains why the multi-site and uni-site models predict similar effectiveness. Under scenario 4, the proportion of susceptibles to genital infection is exceptionally lower than for the uni-site model: natural immunity post-extragenital infection hinders the transmission to genital sites.

Current evidence from the literature seems to lend more support to the assumption of systemic immunity following clearance of genital infection (scenario 3)[27;29;210]. To our knowledge, there is no direct evidence and no literature about the possibility of local immunity against HPV infections (scenario 1 & 2). Yet, both the acquired humoral and cell-mediated immune system could theoretically

have site-specific differences, which could result in greater natural immunity at the site of a previous infection. For example, Tissue-Resident Memory T-cells could be responsible for differential local immunity[211]. On the other hand, vaccination should induce systemic HPV immunity, which is supported by recent studies[164]. Whether systemic immunity also extends to naturally acquired antibodies remains unknown. If this was the case, systemic immunity would be more likely following cervical HPV infection than infection at any other sites, because the rate of seroconversion is the highest following cervical infection and is very low for other sites of infection[27;29]. Thus, the higher rate of seroconversion in women should result in greater protection of women against extragenital HPV infections, and this has been proposed as an explanation for the gender-difference in oral HPV prevalence[47]. However, a protective effect of antibodies on acquisition of extragenital infections has not yet been demonstrated[210;212;213]. Furthermore, prevalence of anal HPV is not lower in women compared to men, but this could be due to a strong correlation in the timing of anal and genital HPV acquisition in women (hence women may acquire anal HPV before acquiring natural immunity).

Our study is the first to calibrate a multi-site model to HPV prevalence to assess differences in predicted effectiveness with traditional uni-site models. To our knowledge, two multi-site models have been published[190;214;215], and none of which has examined the impact of vaccination. In particular, Brouwer et al.[190] have shown that a substantial bias can occur by calibrating a model without autoinoculation if the true model generating the data has autoinoculation. Our results show that models with autoinoculation predict lower effectiveness than models without autoinoculation in some specific contexts (e.g., when the proportion of individuals susceptible to genital infections is similar to the proportion of those susceptible to extragenital infections). However, the effect of autoinoculation was much lower than in the theoretical example presented in Bouwer et al. This may be because Bouwer et al., did not include natural immunity in their models and did not calibrate their model to endemic prevalence of HPV. Hui et al.[215] have shown that pharyngeal and anal infections by gonorrhea can explain the substained transmission to the urethral site in a Men-who-have-Sex-with-Men population in which transmission occurs through oral $\leftrightarrow$ genital, oral $\leftrightarrow$ anal, and anal $\leftrightarrow$ genital contacts. They are able to show that transmission of gonorrhea can be disrupted by preventing only oral $\leftrightarrow$ genital transmission. Unlike the work of Hui et al.[215], we cannot determine from the calibration we performed whether a specific HPV transmission pathway (e.g., genital $\rightarrow$ oral autoinoculation) is essential or important for sustained transmission of HPV infections. This would require further knowledge on the relevant modes of HPV transmissions which could include non-penetrative acts such as kissing or sexual touching.

This study has three main limitations. First, for simplicity, we calibrated our models using probabilities of HPV transmission, while other parameters remained fixed at values extracted from the literature. We examined different assumptions (and values) of natural immunity. Varying the probability of natural immunity affects the proportion of susceptibles to infection at the different sites. We observed that increasing the probability of natural immunity to extragenital infections from 0%

(scenario 1) to 45% (scenario 2) decreased the proportion of susceptibles to extragenital infections and thus decreased predicted effectiveness with the multi-site model. Varying clearance rates also affects the proportion of susceptibles: for a given prevalence of infection, increasing clearance rates increases the proportion of immune individuals and decreases the proportion of susceptibles. Second, we assumed near-symmetrical transmission parameters between women and men. We show in the Supplementary materials that there may be additional dynamics to consider when the prevalences are highly asymmetrical between women and men, but the bounds on the elimination threshold we observed in Figure 1.2 would not change. Finally, we did not include specific sexual acts (e.g., oral sex) in our model, which implies that there is no within-individual correlation in sexual practices.

HPV may be able to infect other sites than the anal, genital and oral canals. For instance, nails are known to harbor HPV DNA and subungual cancers have been attributed to HPV16[216].The inclusion of these other sites of infection in HPV models could affect predictions of vaccination effectiveness against genital infection only if infections at these sites can be transmitted to the genital site (even indirectly) or if they contribute to natural immunity to genital infections. Some of the results presented here can be generalized to any number of sites. Thus, if the simple multi-site HPV model was to include three or more sites of infection (e.g., genital, oral and anal), predicted effectiveness would be in-between effectiveness predicted with two uni-site models of the two sites with the highest and lowest proportions of susceptibles. However, for the heterogeneous multi-site model of objective 1, the minimum predicted effectiveness with the multi-site model can theoretically be lower than the effectiveness predicted with a uni-site model fitted to genital HPV (the site with the lowest proportion of susceptibles) as shown in Figure 1.1B. This phenomenon can be amplified with additional sites (see Supplementary materials A.2.4).

1.7 Conclusions

In conclusion, for the assessment of vaccination effectiveness against genital infections and diseases, multi-site transmission of HPV is important to model if: 1) a significant proportion of genital infections originates from an extragenital site, or 2) extragenital infection contributes significantly to the natural immunity against genital infection. Currently, there is no strong evidence that extragenital infections are a reservoir for genital infections in heterosexual transmission of HPV or that natural immunity following extragenital infections would protect against future genital infections. Hence, the possibility of a strong bias from using a uni-site model to assess vaccination effectiveness against genital HPV16 in women is unlikely given our current understanding of the natural history of HPV infection.

Chapter 2

Impact of changes in sexual behavior on past and future trends of HPV infections and related cancers.

2.1 Résumé

L'objectif de cette étude est de comprendre l'impact des changements de comportements sexuels sur les tendances passées des cancers de l'oropharynx (CO) et du col de l'utérus (CU) reliés au virus du papillome humain (VPH), et de prédire l'incidence future de ces cancers. Nous avons développé un modèle mathématique individus-centré simulant la transmission du VPH16, la progression du VPH16 vers le cancer (CO et CU) et les comportements sexuels des américains nés entre 1850 et 1999. Puis, nous avons simulé l'incidence des CO et CU reliés au VPH16 entre 1862 et 2045. En absence de dépistage du CU, notre modèle prédit une augmentation de 120% dans l'incidence du CU entre 1975 et 2015. De plus, notre modèle prédit une augmentation de 310% dans l'incidence de CO entre 1985 et 2015, ainsi qu'une augmentation de 50% entre 2015 et 2045. Cette augmentation future du CO toucherait des cohortes d'hommes non-vaccinés.

2.2 Abstract

Introduction: Changes in sexual behavior are hypothesized to be the main cause of the substantial rise in oropharyngeal cancers in the past two decades in developed countries. Linear extrapolation of current trends have shown that incidence of HPV-positive oropharyngeal cancer could exceed the number of cervical cancers by 2020. However, these projections are not informed by the possible causes of past increases, such as changes in sexual behavior over time. The aim of this study is to understand the potential impact of changes in sexual behavior on past trends of HPV transmission and to predict future trends of HPV-related cancers.

Methods: We developed an individual-based model of HPV16-transmission using sexual behavior data from three US national surveys and one survey from France. The model reproduces the history of sexual intercourse and oral sex of birth cohorts from 1850 to 1999. We performed simulations of HPV16 transmission, and progression to oropharyngeal and cervical cancers from 1862 to 2045, taking into account changes in sexual behaviour. Model parameters were calibrated to HPV16 prevalence and US incidence of HPV16-related oropharyngeal and cervical cancers. We considered scenarios with/without oral sex. We estimated past relative increase in HPV16-positive cervical cancer (2015 vs 1975), in HPV16-positive oropharyngeal cancer (2015 vs 1985), and predicted future changes (2045 vs 2015) in oropharyngeal cancer among white men. We compared model predictions to observed trends (Surveillance, Epidemiology, and End Results(SEER) data and HPV-typing from Chaturvedi 2011).

Results: In our simulations, cervical cancer is predicted to have increased between 1975 and 2015 by 120%-280%. HPV-related oropharyngeal cancer is predicted to have increased between 1985 and 2015 by 60%-80% in scenarios without oral sex, compared with an increase of 390% (95% CI 330%-550%) observed in the US population. When changes in the practice of oral sex were modelled, we could reproduce greater increase in HPV16-associated oropharyngeal cancer (310%-1500%). Under the best fitting scenario, the incidence of HPV-related oropharyngeal cancer is predicted to increase by about 50% between 2015 and 2045.

Conclusion: Our results have two main implications: 1) vaccination is unlikely to affect predicted increases in oropharyngeal cancer since most birth cohorts affected by these trends are unvaccinated, and 2) we could currently be measuring HPV vaccination effectiveness within an underlying context of increasing HPV-related cancers due to changes in sexual behavior.

2.3 Introduction

The human papillomavirus (HPV) is a sexually transmitted infection and a necessary cause of cervical cancer, as well as a cause involved in the majority of anal (≈90%) and oropharyngeal (≈60%) cancers[7]. In the past three decades, there has been a substantial increase in HPV-related oropharyngeal cancers in most high-income countries (HIC)[140]. This increase in oropharyngeal cancer has affected predominantly white males, in whom the incidence has increased more than 2-fold in the same period[16]. The rise in oropharyngeal cancers has been driven by HPV-positive cancers despite a decline in HPV-negative cancers which are typically caused by smoking and alcohol consumption. These trends have been hypothesized to be the result of increases in sexual transmission of HPV infection due to sexual behavior changes following the 1960s[12;217]. In the US and other HIC, sexual behavior has changed dramatically over the past century. The National Health and Nutrition Examination Survey (NHANES) from the US shows that the median lifetime number of sexual partners has increased from 2.6 in women born in 1940-1949 to 5.3 for those born in 1970-1979 (6.7 to 8.8 for men)[10]. For

the same birth cohorts, age at sexual initiation has decreased by 2 years for women and 1 year for men. Increases in oral and anal sex have also been observed after the "sexual revolution" in the 1960s: among women from France aged 35-39 years old, the proportion who reported ever receiving cunnilingus increased from 55% in 1970 to 93% in 2006[11;12]. However, in the last two decades, there are evidences of sexual behavior having stabilized, and the number of partners may even have decreased among those born after 1980[17;18]. An increase in sexually transmitted HPV infection in cohorts born between 1940 and 1980 would be expected to also have produced an increase in the incidence of cervical cancer (as HPV is the necessary cause). However, the expected increase in cervical cancer may have been mitigated by the progressive increase in screening participation, which was introduced in the 1950s in HIC[218].

Understanding the relation between changes in sexual behavior and trends in HPV-related cancers would permit better forecasting of future cancer incidence, and better predict the impact of interventions (e.g., screening, vaccination). For example, currently, forecasts for HPV-related oropharyngeal cancer incidence are being made using crude linear extrapolation of current trends[16;219]. These forecasts are predicting that the number of HPV-related oropharyngeal cancers will surpass the number of cervical cancers by 2020. Including such projections in economic analysis of HPV vaccination of girls and boys can have an important impact on the potential burden that can be prevented by vaccination and thus the cost-effectiveness of vaccination. Importantly, more than 80% of oropharyngeal cancers occur after 50 years old, while the peak of HPV incidence occurs between ages 20-24 years. Hence, the predicted rise in oropharyngeal cancer in the next two decades will affect unvaccinated cohorts, and the sexual behavior relevant to these cancers has mostly already occurred. There is thus an opportunity to predict future incidence of oropharyngeal cancer based on the evolution of sexual behavior. However, the current linear forecast of oropharyngeal cancer incidence is likely an overestimation of the true future trends because they reflect changes in sexual behavior that occurred in cohorts born in the 1960s, but they do not account for the recent stabilization of sexual behavior in US and other HIC[16]. Similarly, the assessment of the impact of interventions can be biased if changes in sexual behavior are unaccounted for. For example, the reduction of cervical cancer incidence from cervical cancer screening would be underestimated without accounting for the increase in sexual activity in younger birth cohorts[218]. Taking in account these changes in sexual behavior may provide insight on the true impact of screening over the past 50 years.

Transmission-dynamic models are ideal to examine the relationship between changes in sexual behavior and HPV-related cancer incidence. These models integrate key aspects of sexual behavior and natural history of infection to simulate complex transmission dynamics. They can account for non-linear effects of sexual behavior on HPV transmission which would not be possible with non-mechanistic statistical models. Transmission-dynamic models have already been used extensively to simulate HPV transmission (as well as other sexually transmitted infections) and helped inform decisions about the prevention of HPV-related diseases[170;181;220–222].

The objectives of our study were to develop a dynamic model of HPV transmission and related cancers in the US, integrating changes in sexual behavior over the past century in order to 1) understand the potential impact of changes in sexual behavior on trends of HPV transmission and HPV-related oropharyngeal and cervical cancers, and 2) predict future trends of HPV-related oropharyngeal cancers based on the evolution of sexual behavior.

2.4 Methods

2.4.1 Model overview

We developed an individual-based model of HPV transmission. The model simulates white heterosexual men and women living in the US and born between 1850 and 1999. The population was restricted to white men and women because the rise in HPV-related cancers associated with changes in sexual behavior is observed mostly within this population. We modelled exclusively HPV genotype 16 (HPV16) which is responsible for around 55% of cervical cancer[223], and 85% of HPV-positive oropharyngeal cancers[3]. The model has four components reproducing: US demography, sexual behavior, HPV16 transmission, and development of HPV16-related cancers. See the Technical appendix for an in-depth description of the model and methods.

Demography

Each year between 1862 and 1999, new individuals enter the simulated population on their 12^{th} anniversary according to a pre-determined number of births per year and gender, on historical US census data[224].

Sexual behavior

We determined the sexual activity of each individual between the ages of 12 and 44 years. This period corresponds to the peak of sexual activity[225] and of acquisition of HPV infections[41] from which HPV-related cancers may develop years or decades later. To represent sexual behavior, three phases were performed: 1) sexual history of each individual in the model is determined, 2) sexual partnerships are formed by matching partners, and 3) whether oral sex is practiced in during the partnership is determined.

- **Phase 1. Determining sexual history profiles.** Sexual behavior is determined by the dates of occurrence of every cohabitation and non-cohabitation partnership, the length of cohabitation partnerships, and the age of every sexual partner. The parameters used include the mean and variance of the lifetime number of partnerships, the rate of sexual activity initiation, the rates of cohabitation initiation and separation, the probability of a non-cohabitation partnership occur-

ring in a given month, the mean and variance of the age of a sexual partner. These parameters are age- and year of birth- specific and were obtained by fitting statistical age-cohort models to data from three different surveys: the National Survey of Family Growth (NSFG), the National Health and Social Life Surveys (NHSLS), and the General Social Survey (GSS). The total population of these surveys counted more than 80 000 men and women born between 1900 and 1999. Using the fitted distributions of the parameters, we determined successively: 1) the age at first sex, 2) the dates of start and end of every cohabitation involving a sexual partner, 3) the lifetime number of sexual partners, 4) the dates of every sexual partner not involving cohabitation (instantaneous partnerships), and 5) the age of every sexual partner. Importantly, every step depended on the previous steps, so that for example the age of sexual partners depends on the lifetime number of sexual partners. Figure 2.1 illustrates the step-wise process performed to build the sexual history profile of a simulated individual.

- **Phase 2. Matching of sexual partnerships.** Once the sexual activity profiles are completed for each individual in the model over time, we determined the identity of each partner by matching the sexual partnerships of men and women on the basis of the date, length of partnerships, and ages of partners.
  When matching the partnerships of men and women, there is a discrepancy in the reported number of sexual partners inherited from the sexual behavior surveys data in which men generally report many more partners than women: men report 14.14 partners in their lifetime compared with 7.12 for women in the UK [116;117;119]. To correct this discrepancy, we considered three scenarios: 1) *High-risk group*: a very small high-risk group (e.g., sex workers) of women with high number of partners (>50) is added to the population to compensate for the excess partnerships among men, 2) *Report bias men & women*: the number of partnerships among men is lowered and the number of partnerships in women is increased, and 3) *Report bias men only*: the number of partnerships among men is lowered. These scenarios were based on plausible explanations for the discrepancy: under-sampling and under-reporting (e.g., right-censoring of the number of partners) of women with high number of partners (e.g., sex workers) [119], men overestimating their number of partners (e.g., rounding up when estimating) [116], and women underestimating their number of partners (e.g., pressure of social norms) [117].

- **Phase 3. Determining the practice of oral sex.** We reproduced the practice of oral sex in men only by using data from NSFG 2011-2015 and three surveys from France [12]. Every man in the model had a probability of performing oral sex with a sexual partner that depends on the year of birth and the type of sexual partnership (cohabitation or not). We fitted these probabilities to reproduce 1) the proportion of men who have ever performed oral sex on a sexual partner, 2) the proportion of women who have ever had oral sex performed on them by year of birth. In all scenarios, we assumed that men did not perform oral sex with members of the high-risk group of women.

HPV16 transmission

HPV16 transmission is modelled using two probabilities: 1) the probability of acquiring a genital infection within a partnership with someone infected at the genital site, 2) the probability of acquiring an oral infection within a partnership with someone infected at the genital site and on whom oral sex is practiced. Infection acquisition is determined upon partnership formation (i.e., instantaneous partnerships). Once acquired, infection may clear over time. Upon clearance, either the individual develops lifelong natural immunity or becomes susceptible again.

We considered two different scenarios regarding transmission: 1) *uni-site*, 2) *multi-site with oral sex*. The goal was to assess whether the inclusion of the practice of oral sex and oral infections were necessary to explain trends in oropharyngeal cancer. Furthermore, these scenarios reflect structural uncertainty regarding how HPV is transmitted. In the *uni-site* scenario, HPV16 is modelled as a uni-site infection (genital infection only) in men and women. In the *multi-site* with oral sex scenario, oral infection is included as a separate infection from the genital infection. In this scenario, oral HPV16 infection can only be acquired through oral sex with a partner infected at the genital site. In all scenarios, we fitted the probabilities of transmission per partnership to genital and oral HPV16 prevalence targets based on the literature[41;47] (see Table 2.1). Calibration was performed by the bisection algorithm, testing values between 0 and 1 for the probabilities of transmission. Table 1 describes the different scenarios, the parameters, and the calibration targets.

HPV16-related cancers

Infected individuals can progress toward either oropharyngeal or cervical cancer (oropharyngeal cancer is only modelled in men). The progression to cancer is a function of two parameters that are gender- and site-specific: 1) the proportion of newly acquired infection that progress toward cancer, 2) the dwell time from infection to cancer for infections that will progress to cancer. Dwell time was assumed to be normally distributed. The parameters of progression to cancer were calibrated to reproduce the observed age-specific incidence of HPV16-positive cervical and oropharyngeal cancers in the US by minimizing the mean squared error. The values of the fitted parameters for cancers and HPV transmission can be found in Appendix C section 10. The data were taken from Surveillance, Epidemiology, and End Results (SEER) database. For cervical cancer, we used the incidence from the earliest years of the SEER database (1973-1975) to minimize the impact of cancer screening on incidence. For oropharyngeal cancer, we used the incidence between the years 1984 and 1985, which corresponds to the beginning of the increase in incidence as well as the earliest period of available data on HPV16 positivity. After extracting the overall incidences of cervical and oropharyngeal cancer for these years, we applied the proportion of cancer who are HPV16-positive squamous cell carcinomas to the overall incidence. For cervical cancer that proportion was 55% based on a meta-analysis of Clifford et al.[223]. For oropharyngeal cancer, that proportion was 40% between 1984-1985 based on the study of Chaturvedi et al.[16]. In the rest of the manuscript and when there is no possible ambigu-

ity with adenocarcinomas, we refer to squamous cell carcinoma of the cervix and the oropharynx as cervical and oropharyngeal cancer.

2.4.2 Model predictions

We ran 20 simulations of HPV transmission for each combination of the three sexual behavior scenarios (*High-risk group*, *Report bias men & women*, *Report bias men only*) and of the two transmission scenarios (*Uni-site*, *multi-site with oral sex*). Predictions are the median incidence of HPV16-related oropharyngeal and cervical cancers over time between 1915 and 2015.

To assess the potential increase in cervical cancer that would have occurred if there had been no screening introduced in the 1950s, we estimated the relative increase in cervical cancer incidence between 1975 and 2015. We compared these results with two historical trends observed in the SEER data: 1) the relative decrease since 1975 in incidence of HPV16-positive squamous cell carcinoma of the cervix, and 2) the relative increase since 1975 in adenocarcinomas of the cervix. The sensitivity of pap tests to detect pre-cancerous adenocarcinomas is much less than for squamous cell carcinomas[226;227]. Thus, trends in adenocarcinomas can serve as a pseudo-counterfactual of the incidence of cervical cancer in the absence of screening. Comparing the increase in cervical cancer in our simulations with the historical increase in adenocarcinomas is used as a way to assess the face validity of our results from each scenario.

Similarly, for oropharyngeal cancer, we assessed which of the sexual behaviour and HPV transmission scenarios could reproduce the historical trends in HPV16-positive oropharyngeal cancer between 1985 and 2015 estimated from the SEER data. In particular, we were interested in determining if including the practice of oral sex (scenario *multi-site with oral sex*) produced more realistic trends in oropharyngeal cancer incidence. HPV16-positivity in cervical and oropharyngeal cancers trends were estimated from the same studies[16;223].

We also predicted incidence of HPV16-positive oropharyngeal cancer up to 2045 among white men aged more than 50 years old. This age stratification was done to make sure we only estimated incidence among men that were included in the model population (i.e., born before 1999).

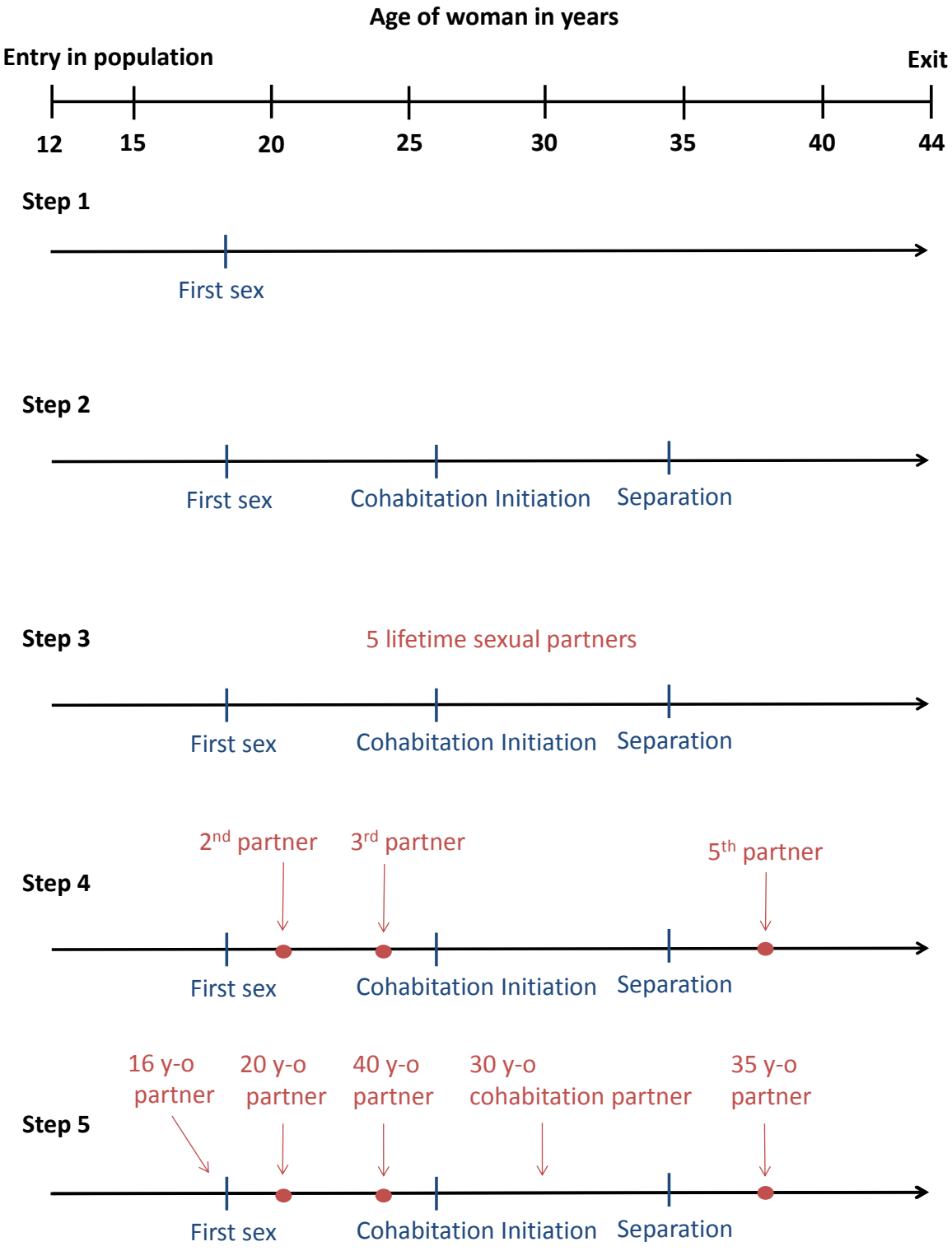


Figure 2.1 – **Illustrative example of how a sexual history profile is determined.** The woman in this example has 5 lifetime sexual partners: 4 non-cohabiting partners and 1 cohabitation partner. The sexual history profile characteristics were estimated in the following order: Step 1) the age at first sex, Step 2) the dates of start and end of every cohabitation involving a sexual partner (possibly more than one), Step 3) the lifetime number of sexual partners, Step 4) the dates of every sexual partner not involving cohabitation (instantaneous partnerships; shown in red), and Step 5) the age of every sexual partner.

Table 2.1 – **Model scenarios and calibration**

	Scenario uni-site	**Scenario multi-site with oral sex**
Calibration target: HPV16 prevalence	**Genital** 9% among 20- to 30-years old in 2015[41]	**Genital** 9% among 20- to 30-years old in 2015[41] **Oral** 2% among 20- to 30-years old men in 2015[47]
Varying parameters	Probabilities of transmission per partnership: • Genital → Genital	Probabilities of transmission per partnership: • Genital → Genital, • Genital → Oral for partnerships involving oral sex
Fixed parameters	Average duration of genital infection: • 1.5 years[206] Probability of developing natural immunity following clearance of genital infection[207]: • 35% (women) • 10% (men)	Same values as in scenario *uni-site* for genital infections. Average duration of oral infection: • 1.5 years (based on cervical HPV[206]) Probability of developing natural immunity following clearance of oral infection: • 0%

2.5 Results

Table 2.2 shows the characteristics related to sexual behavior and its evolution across birth cohorts of the model population. From 1900 to 1985, the median lifetime number of sexual partners and the median age at initiation of first cohabitation went up, and the median age at first sexual intercourse went down for both men and women. These changes in sexual behavior simulated in the model population reproduced the changes in the US population as observed in the data from NSFG, NHSLS, and GSS (see Technical appendix C for figures of model fit to the US population data).

Table 2.2 – **Changes in sexual behavior by birth cohorts as reproduced in the model population.**

	Women				
	Birth year				
	1900-20 Med $(25^{th}-75^{th})$	1921-40 Med $(25^{th}-75^{th})$	1941-60 Med $(25^{th}-75^{th})$	1961-70 Med $(25^{th}-75^{th})$	1971-85 Med $(25^{th}-75^{th})$
Lifetime number of sexual partners* for the 3 scenarios:					
High-risk group	1 (1-1)	1 (1-2)	2 (1-4)	4 (1-7)	5 (2-9)
Report bias men & women	1 (1-2)	1.5 (1-2)	3 (1-6)	5 (2-9)	5 (2-11)
Report bias men	1 (1-1)	1 (1-2)	2 (1-4)	4 (1-7)	5 (2-9)
Age at first sexual intercourse (years)	19.2 (17.7-21.3)	18.9 (17.5-20.9)	18.5 (17.1-20.5)	17.8 (16.2-19.5)	17.2 (15.6-19.2)
Age at initiation of first cohabitation (years)	20.1 (18.3-22.8)	20.0 (18.4-22.8)	20.7 (18.7-23.6)	21.8 (19.2-24.9)	21.8 (19.3-25.0)
	Men				
	Birth year				
	1900-20 Med $(25^{th}-75^{th})$	1921-40 Med $(25^{th}-75^{th})$	1941-60 Med $(25^{th}-75^{th})$	1961-70 Med $(25^{th}-75^{th})$	1971-85 Med $(25^{th}-75^{th})$
Lifetime number of sexual partners* for the 3 scenarios:					
High-risk group	1 (1-3)	2 (1-6)	5 (2-10)	5 (3-11)	6 (3-13)
Report bias men & women	1 (1-3)	2 (1-5)	4 (2-10)	5 (3-11)	6 (3-13)
Report bias men	1 (1-2)	2 (1-5)	4 (2-10)	5 (2-9)	5 (3-11)
Age at first sexual intercourse (years)	20.5 (17.0-22.9)	19.2 (16.8-21.8)	17.9 (16.0-20.4)	17.4 (15.6-19.8)	17.0 (15.3-19.4)
Age at initiation of first cohabitation (years)	22.8 (20.7-27.1)	22.6 (20.6-27.0)	23.0 (20.7-27.0)	24.5 (21.3-30.0)	24.2 (21.2-29.0)

Lifetime number at 30 years old, Med= median; $25^{th}-75^{th}$= 25^{th} and 75^{th} percentile.

*

Cervical cancer

Figure 2.2 shows the observed and predicted incidence of HPV16-related cervical cancer over time. In the US population, incidence of cervical cancer has declined by 53% (95% CI 50%-55%) between 1975 and 2015. However, our model predicts that, without screening or vaccination, the incidence of cervical cancer would have increased by 120% to 280% during the same period depending on the model assumptions. The predicted increase was the lowest for the scenario *Report bias men & women* (increase of 120%) while the two other scenarios had similar predictions.

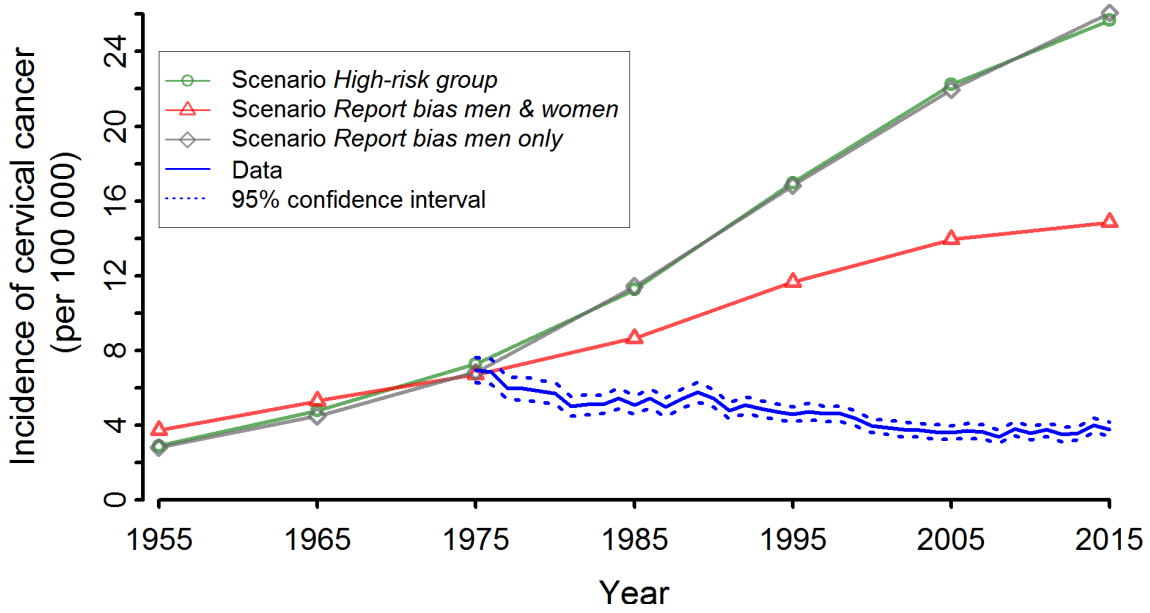


Figure 2.2 – **Incidence of HPV16-positive squamous cell carcinoma of the cervix by year of diagnostic: model predictions vs observed data from SEER.** For each scenario and each 10-year period, the median incidence of the 20 model simulations is retained.

Figure 2.3 shows the predicted incidence of cervical cancer compared with the observed incidence of adenocarcinomas of the cervix over time. In the US population, the incidence of adenocarcinoma of the cervix increased by 80% (95% CI 60%-100%) from 1975 to 2015. This increase is lower than the increase in incidence predicted in scenarios *High-risk group* and *Report bias men only* (250% and 280% respectively), but comparable, albeit slightly less, to the increase observed in scenario *Report bias men & women* (120%).

Oropharyngeal cancer

Figure 2.4 shows the predicted and observed incidence of HPV16-related oropharyngeal cancer over time. In the US population, according to our estimation of HPV16-positivity, incidence of HPV16-related oropharyngeal cancer has increased by 390% (95% CI 330%-550%) between 1985 and 2015. Our model predictions varied greatly depending on whether or not the practice of oral sex was included. In scenario *uni-site* without the practice of oral sex, our model predicts relatively small in-

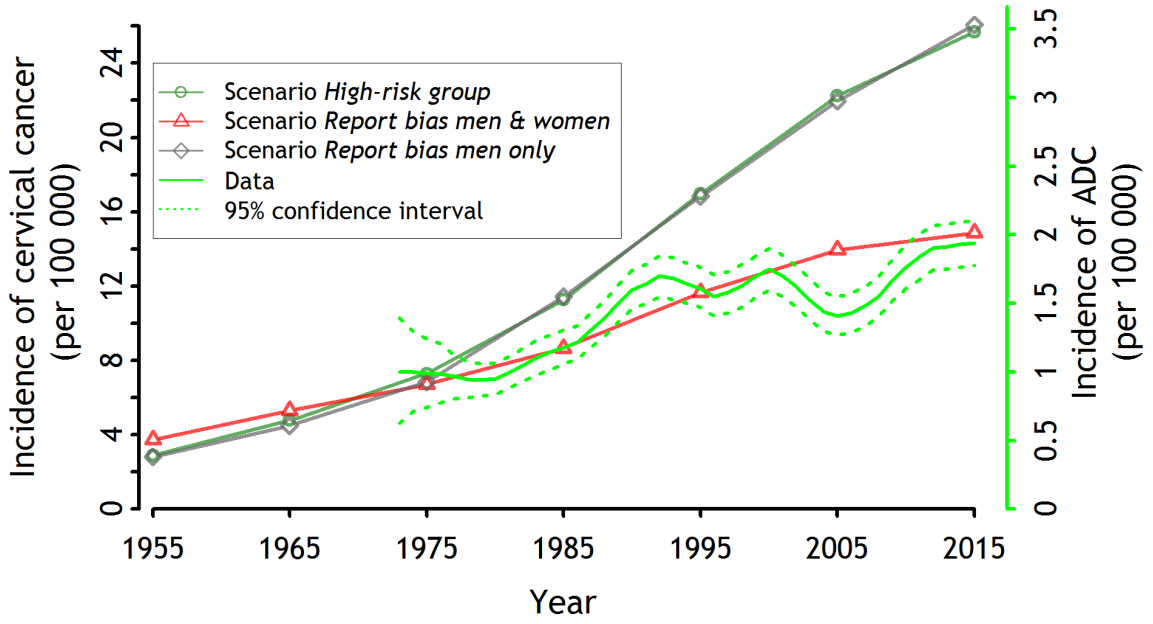


Figure 2.3 – **Incidence of adenocarcinomas of the cervix by year of diagnostic: model predictions of incidence of HPV16-positive squamous cell carcinoma of the cervix vs observed adenocarcinoma incidence from SEER incidence data.** For each scenario and each 10-year period, the median incidence of the 20 model simulations is retained.

crease in incidence of oropharyngeal cancer between 60% and 80% depending on the other model assumptions (see Figure 2.4A). In scenario *multi-site with oral sex*, our model predicts steep increase in incidence between 310% and 1500% (see Figure 2.4B). As with cervical cancer, the predicted increase is the lowest when assuming under-reporting in women and over-reporting in men (*Report bias men & women*).

Projections from 2015 to 2045

Figure 2.5 shows the projection of HPV16-positive oropharyngeal cancer under the best-fitting scenario: *Report bias men & women* with *multi-site with oral sex*. Incidence of oropharyngeal cancer is predicted to increase by 50% from 2015 to 2045.

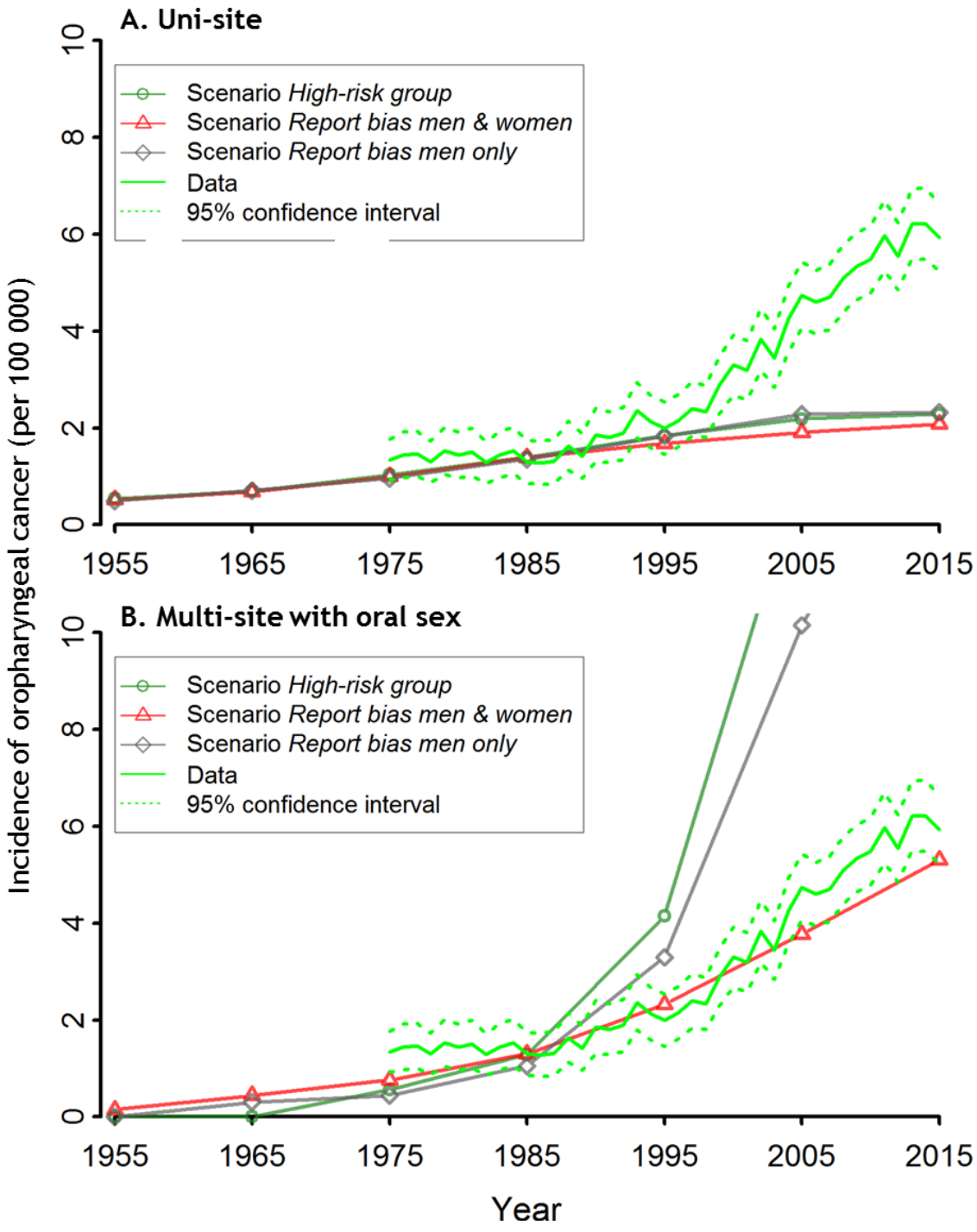


Figure 2.4 – **Incidence of HPV16-positive oropharyngeal cancer in men by year of diagnostic: model predictions from vs observed incidence data from SEER. Model simulations are from scenarios A) *uni-site*, B) *multi-site with oral sex*.** For each scenario and each 10-year period, the median incidence of the 20 model simulations is retained.

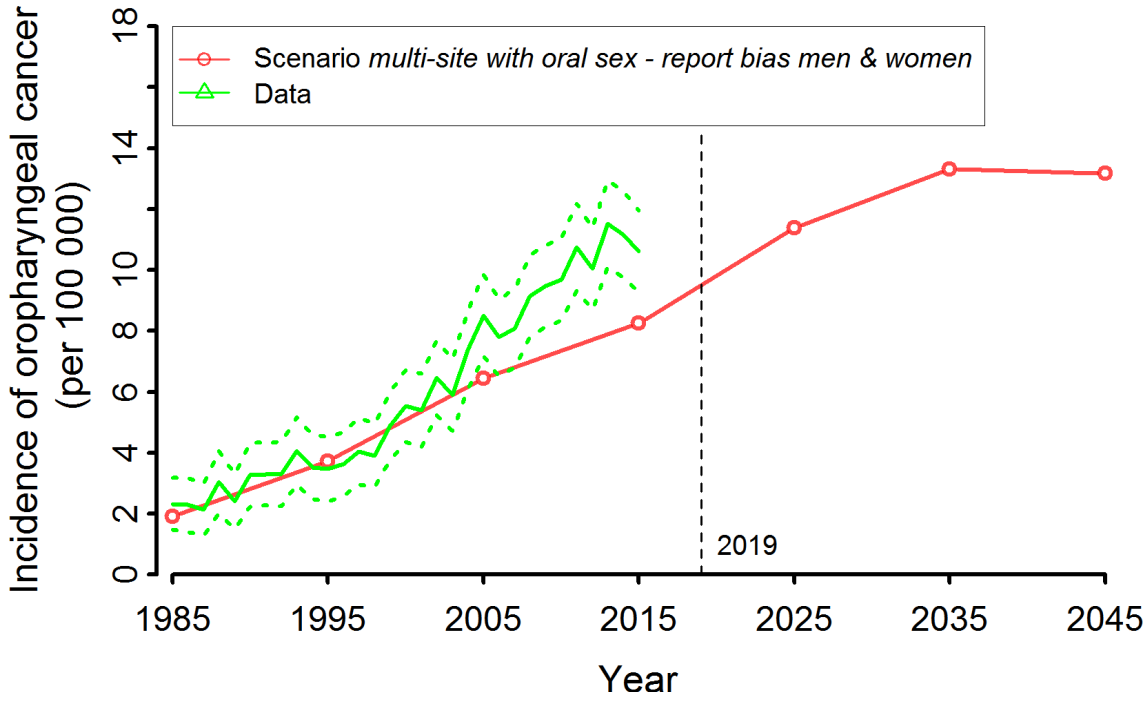


Figure 2.5 – **Incidence of HPV16 positive oropharyngeal cancer in men aged more than 50 years old by year of diagnostic: model predictions from scenario *multi-site with oral sex - report bias men & women*.** For each scenario and each 10-year period, the median incidence of the 20 model simulations is retained.

2.6 Discussion

We reproduced, in a HPV transmission-dynamic model, the changes in sexual activity over the 20^{th} century as observed in three national surveys of the US population: NSFG, NHSLS, and the GSS. We observed a progressive increase in the lifetime number of sexual intercourse partners reported by men and women over the last century, with trends in men stabilizing before trends in women. The relative increase in partners was larger in women compared to men: the number of partners in women doubled in the span of 20 years, while the number of partners only increased by 20%-50% in men. Our simulations suggest that the increase in lifetime number of sexual intercourse partners would have resulted in at least a doubling of the incidence of cervical cancer over the past 40 years, in the absence of cervical screening. However, these changes in the practice of sexual intercourse alone produced a small increase in oropharyngeal cancer incidence among men, unlike the substantial increase observed in US data. Changes in the practice of oral sex were necessary to reproduce US data of oropharyngeal cancer incidence in the model population. Finally, our simulations predict that, if oral sex is the main mode of oral HPV transmission, incidence of oropharyngeal cancer could increase until 2045.

Our study supports an increase in HPV transmission for cohorts born after the 1940s. There is little direct evidence of this increase since HPV was isolated only in 1983. Yet, studies conducted in Finland and Sweden have shown an increase in HPV16 seroprevalence in pregnant women between 1980 and 2000[228;229]. Chaturvedi et al.[16], have also shown a major increase in HPV-positivity of oropharyngeal cancer specimens during the same period. As for indirect evidence, in a previous Age-Period-Cohort analysis by Vaccarella et al.,[218] an increase in risk of cervical cancer across birth cohorts was estimated but masked by the effect of screening. This cohort effect represented an incidence rate ratio of cervical cancer of around 1.4 between 1940 and 1980. In our study, we obtained a larger incidence rate ratio (2.2-3.8) due to cohort effects (i.e., increase in sexual activity). The difference between our estimates and the one from Vaccarella et al. may be due to factors we did not consider in our simulations (e.g., condom usage, contraceptive use, smoking).

As our results show, the potential reporting biases in sexual behavior surveys can have important consequences on predicted trends of HPV-related cancers. According to our results, the scenario *Report bias men & women* is the most realistic of the three scenarios. Indeed, we obtained a lower incidence of cervical and oropharyngeal cancers and a better fit to the US data under this scenario. When correcting for the under-reporting in women, the lifetime number of sexual partners of women increases closer to the lifetime number of sexual partners of men. Hence, the relative increase in lifetime number of sexual partners of women decreases, which also results in a lesser increase in genital HPV transmission throughout the 20^{th} century. Since oral HPV infection in men is acquired through genital HPV infection in women, correcting under-reporting in women also attenuates the predicted trends in oral HPV infection.

Our results are compatible with an important role of oral sex in the transmission of oral HPV infections

and in the trends of HPV-related oropharyngeal cancer. In this study, changes in sexual intercourse alone could not explain the trends in oropharyngeal cancer. However, the practice of oral sex has increased over the 20^{th} century independently of the number of sexual partners. Indeed, the proportion of women that ever received cunnilingus almost doubled in last three decades based on empirical data from France (see C.11), and this increase cannot only be explained by the increase in lifetime number of partners. Furthermore, in the model population, oral sex is not performed with women from the high-risk group, and the proportion of partnerships with women from the high-risk group has declined over time (because the discrepancy in number of partners between men and women has declined over time). Hence, part of the increase in the practice of oral sex is mediated by the decrease in the proportion of partnerships that are formed with the high-risk group. Some empirical evidences support this: in Gagnon and Simon 1987[11], young men born in the 1940s reported twice as much experience with cunnilingus and less experience with fellatio compared with young men born in the beginning of the century. The authors note that these observations "suggest a reduction in excess experience of fellatio by males from commercial contacts". Hence, in the early 20^{th} century, males had experience of fellatio (and probably sexual intercourse) with sex workers, but may have not performed cunnilingus as often on them. These data on the practice of oral sex, along with the data from France and the NSFG, still leave a significant amount of uncertainty regarding how the practice of oral sex has changed over time in the US population.

There are two main implications to our results. Firstly, we could currently be measuring HPV vaccination effectiveness within an underlying context of increasing HPV-related cervical cancer due to changes in sexual behavior. In such situation, the expected decrease in cervical cancer due to vaccination could be mitigated by the background trend. Hence, without accounting for changes in sexual behavior, the true impact of vaccination could be underestimated in surveillance study or in mathematical model analyses. Secondly, vaccination is unlikely to affect predicted increases in oropharyngeal cancer in the next three decades since most birth cohorts affected by these trends are unvaccinated. Thus, the predicted increase in oropharyngeal cancer could represent a transient but significant population health issue even in countries with gender-neutral vaccination, and could increase the cost-effectiveness of vaccinating boys in countries with only female vaccination.

Our methodology has advantages over the traditional Age-Period-Cohort approach used to understand trends in HPV-related cancers[218;219]. These types of models require the use of statistical assumptions that cannot readily be justified or translated in meaningful terms. Predictions with these models are also based on the period- and cohort- effects which simply reflect the past trends, instead of being based on the evolution of the causal variables. On the other hand, mathematical modelling relies on data and assumptions that relate to the causal variables of the system: namely the parameters of natural history of infection or of sexual behavior. This allowed us to produce predictions based on the evolution of sexual behavior. Another strength of our approach is to have used data on cohabitation history to reproduce the sexual activity of older birth cohorts. In fact, it poses a challenge to cor-

rectly extrapolate the timing of sexual partnerships for older birth cohorts because these partnerships occurred decades before national surveys on sexual behavior were available. However, we were able to do so by using the relation between cohabitation history and timing of sexual partners, which was investigated in NHSLS.

The main limit of this study is the validity of the data on sexual behavior. Sexual behavior surveys are highly sensitive to reporting bias, and such bias may vary according to cultural norms. Hence, the reporting bias may impact the trends over time in sexual activity in men and women. We expect that under-reporting in women to be more frequent in older birth cohorts compared to younger birth cohorts. We have assessed the potential impact of this bias by assuming under-reporting in women and over-reporting in men in three different scenarios. Another limit is that some parameters of sexual activity and their evolution throughout the 20^{th} century were unknown: the timing of non-cohabiting sexual partnerships, the proportion of partnerships on whom oral sex was performed, the occurrence of sexual partners with whom only oral sex was performed, the sexual mixing between men and women with respect to level of sexual activity. Hence, we had to make assumptions to fill in the missing information. For instance, we omitted partnerships on whom only oral sex was performed and we assumed random mixing between levels of sexual activity.

2.7 Conclusion

Our results suggest that cervical screening prevented a sharp increase in cervical cancer in the past decades. Furthermore, the increase in oral sex could be the main factor behind the increasing trends in HPV-related oropharyngeal cancer. If so, the incidence of oropharyngeal cancer could continue to increase among unvaccinated males in the next three decades. Assessment of HPV vaccination effectiveness should account for the evolving burden of HPV-related cancers due to changes in sexual behavior in order to avoid an underestimation bias.

Chapter 3

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

3.1 Résumé

Dans l'évaluation de l'effet causal de facteurs de risque d'infections transmissibles sexuellement et par le sang, de la confondance persistant après ajustement pour les caractéristiques individuelles peut se produire sous deux conditions : (C1) assortativité sexuelle relatif au facteur de risque examiné et (C2) l'activité sexuelle est associée au facteur de risque. Notre objectif est d'illustrer l'impact potentiel de ce biais d'assortativité en prenant comme exemple l'association tabac - VPH. Nous avons développé un modèle de transmission dynamique de VPH à partir duquel nous avons simulé une étude transversale évaluant l'association tabac - VPH. Pour le scénario de base, on a supposé (1) aucun effet du tabac sur le VPH, et (2) les conditions C1-C2. Le biais d'assortativité a causé une surestimation du rapport de cotes (RC) même après l'ajustement complet pour les caractéristiques individuelles (RC ajusté de 1,51 au lieu de 1,00). L'ajustement pour les caractéristiques des partenaires est nécessaire pour mitiger le biais d'assortativité.

3.2 Abstract

Introduction: 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.

Methods: 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).

Results: 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.

Conclusion: Conditions C1 and C2 can cause bias in studies assessing the causal effect of risk factors on acquisition or duration of STIs, as in the case of the effect of smoking on HPV prevalence. Adjustment for characteristics of partners is needed to mitigate the *assortativity bias*.

3.3 Introduction

Contact networks of individuals affect their exposure to infectious contacts so are a crucial determinant of infection risk[230]. 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[231]. 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 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[232;233]. 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 behavior (e.g., smoking status) as themselves[19;234–236].

In STI studies, *assortativity bias* can occur if two conditions are met: (C1) partner choice is assorta-

tive 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 the *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)[235;236], and smokers have higher average level of sexual activity (C2)[237;238].

Figure 3.1 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 3.1, 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 relation is confounded by sexual activity of the subjects (Figure 3.1 right panel, dotted arrows) and subjects' partners (Figure 3.1 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 conditions 1-2 are met for many common risk factors of STIs, including age, race and socio-economic 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.

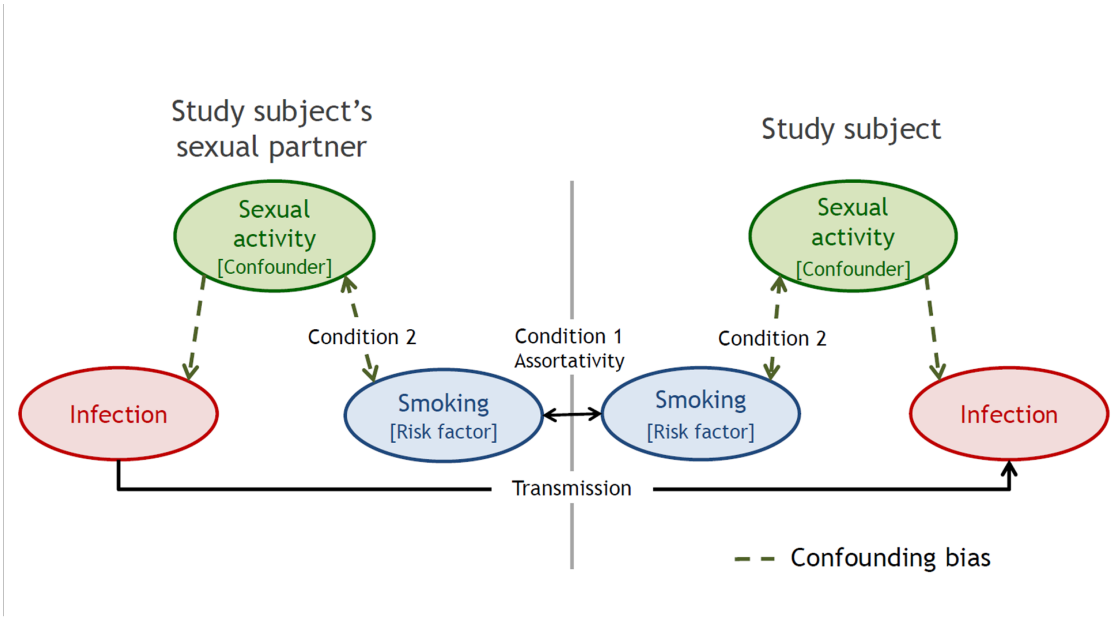


Figure 3.1 – **Illustrated example of the *assortativity bias* in the association between smoking and a sexually transmitted infection**. Directed arrows represent causal link and double-headed arrows represent statistical link (the distribution of one variable changes conditional on the other variable). 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.

Mathematical modeling has been used to understand potential biases in epidemiological studies of STIs. With modeling, 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 behavior, transmission, and natural history[239–242].

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[243]. 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[47;233;237;244;245], and these associations have been shown to follow monotone dose-response relationships[237]. However, the possibility of *assortativity bias* was not addressed in these studies.

The specific objectives of this paper are to use mathematical modeling: 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 behavioral parameters, for generalization of results.

3.4 Material and methods

3.4.1 Mathematical model

We developed a deterministic transmission dynamic model of HPV infection (see Supplementary material for a list of the model's equations). The model population is heterosexual, open and stable. For the base-case scenario, we modeled HPV16 infection, which is the most prevalent and oncogenic type. The simulated population is stratified for the two behavioral 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 model population, representing the years of higher sexual active life (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 (smokers only form partnerships with smokers). We included assortativity according to sexual activity as it is a key feature of sexual networks[246;247]. Based on empirical evidence, we assume that the two conditions for the *assortativity bias* are met: C1) sexual mixing is assortative for smoking[235;236], and C2) smokers are more sexually active than non-smokers[237;238].

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*.

3.4.2 Parameterization

Model parameter values and references are presented in the Appendices (see Table D.1 in Supplementary material). We used biological parameter values estimated in prior modeling work[178;186;248], and estimated the proportion of smokers in each sexual activity class from an epidemiological study[249]. Although studies suggest that sexual mixing by level of sexual activity and smoking status is assortative[235;236;250;251], no empirical estimates of assortativity 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=random, 1.0=complete assortativity), using equations presented in the Appendices (see Supplementary material). We performed extensive sensitivity analysis on mixing parameters given their uncertainty.

3.4.3 Experimental design & 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) 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 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 biologic 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 Prevalence/(1-Prevalence)=Incidence x Duration.

3.4.4 Sensitivity analyses

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 infection.

3.5 Results

3.5.1 Base-case

Table 3.1 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 3.1). 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 lower for those with greater sexual activity (Table 3.1, Figures 3.2 and 3.3). This is because highly sexually active subjects will likely have highly sexually active partners (assortativity by sexual activity), irrespective of smoking status.

3.5.2 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 3.2A). This is because a stronger association causes greater confounding by sexual activity (increased impact of Condition 2, Figure 3.1).

Assortativity by smoking status. Greater assortativity according to smoking status results in a steep

Table 3.1 – **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

increase in the ORs of HPV infection comparing smokers with non-smokers (Figure 3.2B). 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 3.1).

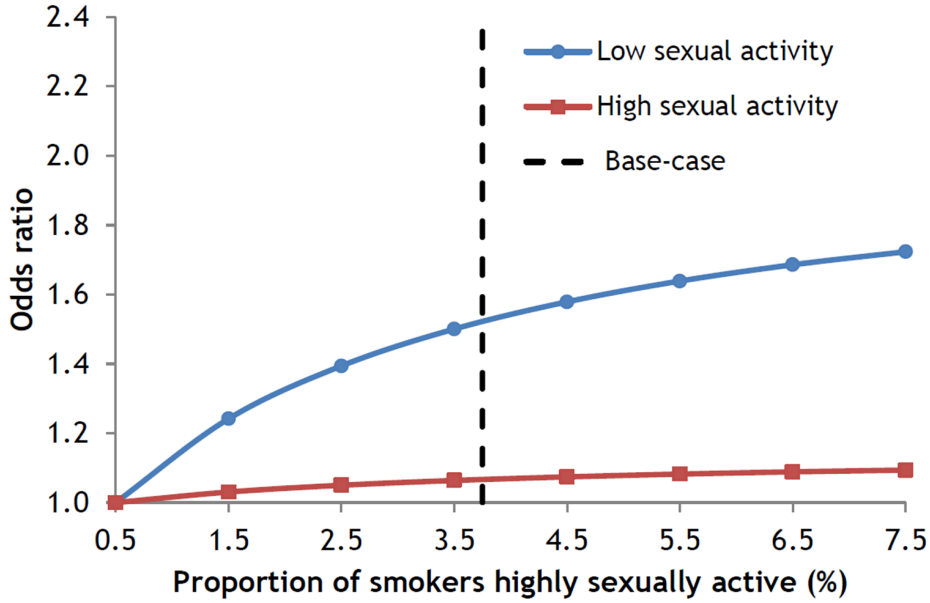
Assortativity by sexual activity. The ORs of HPV infection decrease with greater assortativity between individuals of the same sexual activity class (Figure 2C). As assortativity by sexual activity increases, the sexual activity of the study subject becomes a better proxy of his/her partners' sexual activity. 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 adjustment for sexual activity.

3.5.3 Impact of biological factors

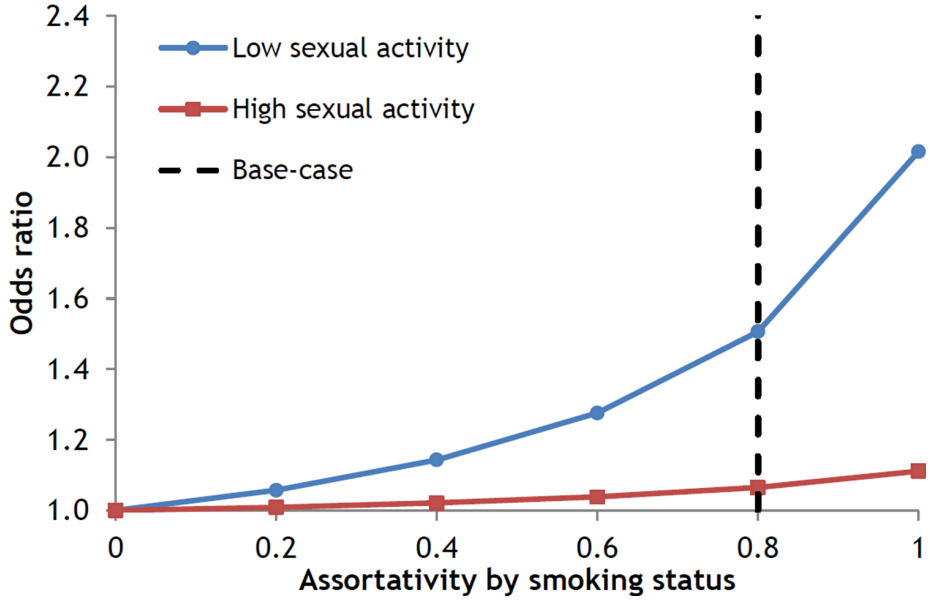
Transmission probability & duration of infectiousness. The model shows that the magnitude of the *assortativity bias* is highly sensitive to the transmission probability or duration of infection (Figure 3.3A- 3.3B). The ORs of HPV infection among smokers compared with non-smokers, stratified by sexual activity of study subjects, decrease 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.

Natural immunity. The probability of developing natural immunity has little impact on the magnitude of the *assortativity bias* (Figure 3.3C). Lower natural immunity has the same relative impact on HPV prevalence among both smokers and non-smokers.

A)


2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
Odds ratio
Low sexual activity
High sexual activity
Base-case
0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
Proportion of smokers highly sexually active (%)

B)


2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
Odds ratio
Low sexual activity
High sexual activity
Base-case
0
0.2
0.4
0.6
0.8
1
Assortativity by smoking status

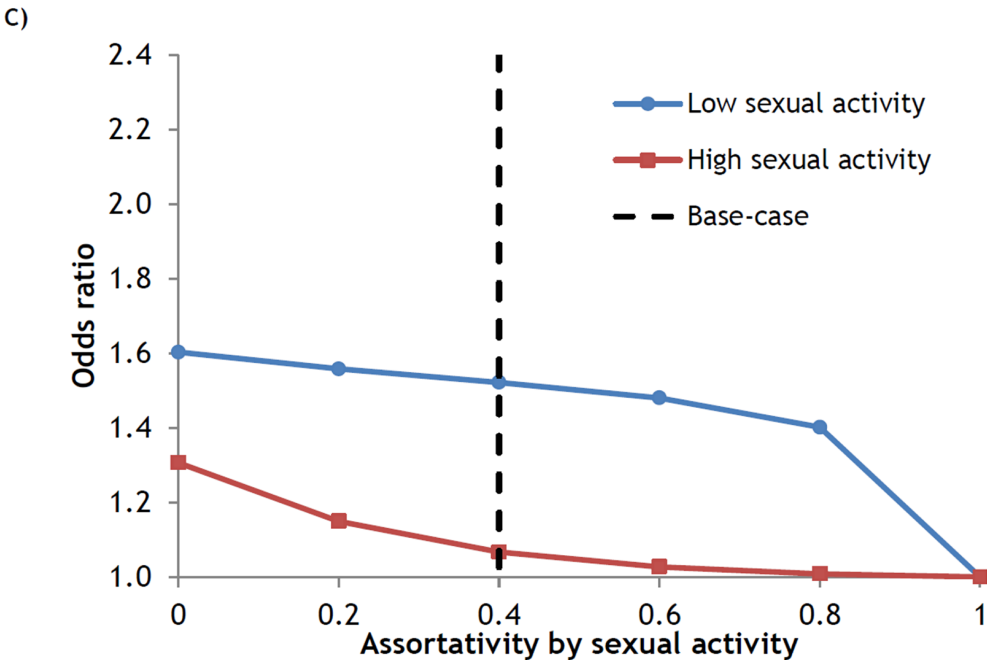
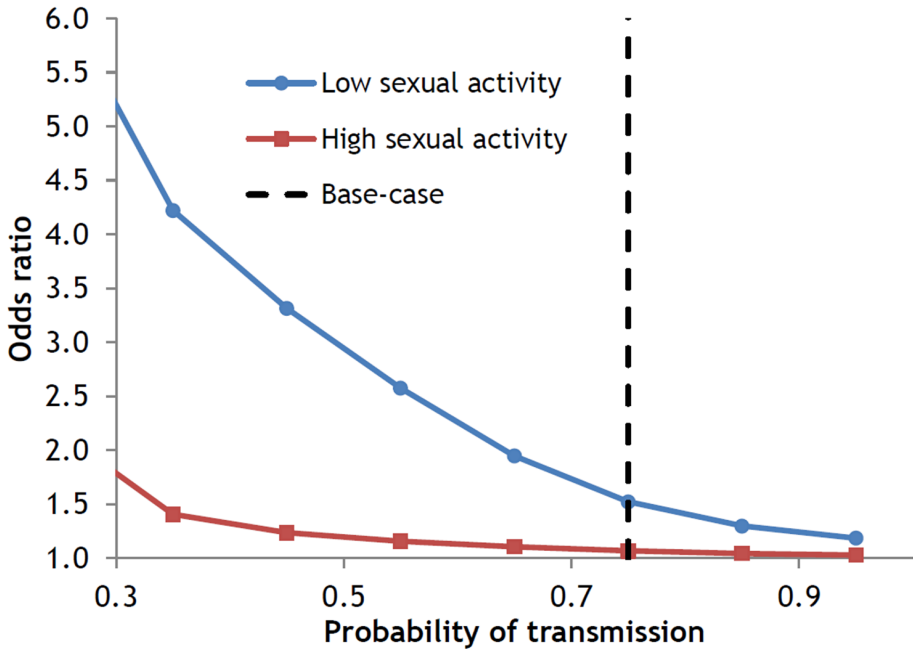
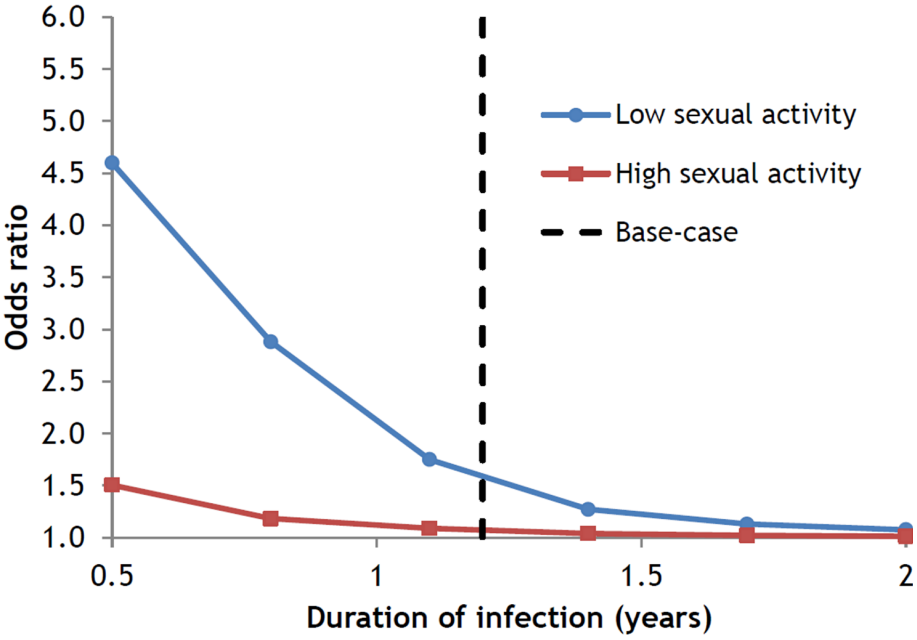


Figure 3.2 – **Impact of behavioral parameters on the *assortativity bias*.** Univariate sensitivity analysis of the odds ratios of prevalence between smokers and non-smokers with one parameter varying: A) Proportion of smokers that are highly sexually active, B) Assortativity by smoking status, and 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.

A)


6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Odds ratio
Low sexual activity
High sexual activity
Base-case
0.3
0.5
0.7
0.9
Probability of transmission

B)


6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Odds ratio
Low sexual activity
High sexual activity
Base-case
0.5
1
1.5
2
Duration of infection (years)

C)

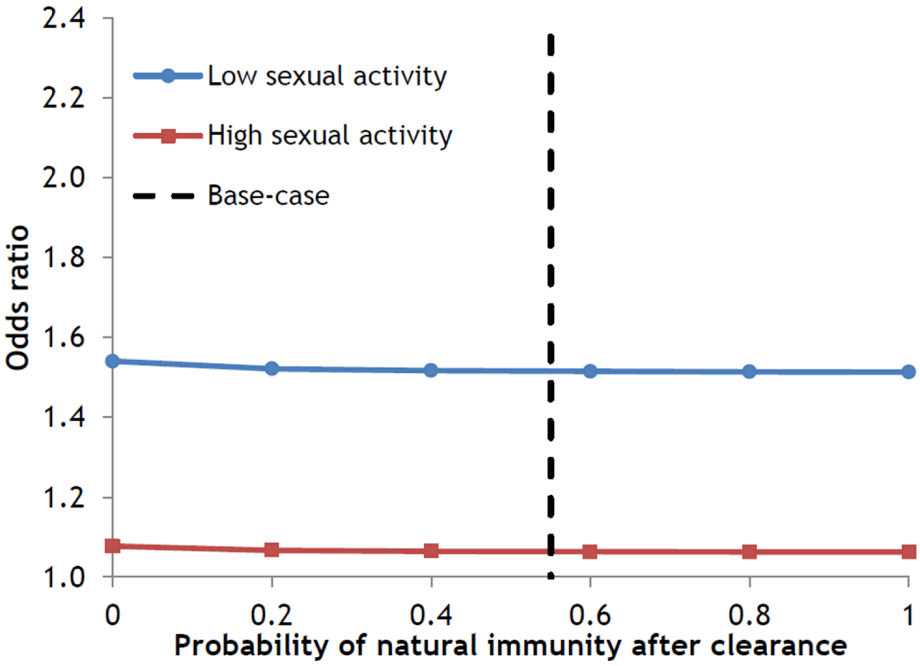


Figure 3.3 – **Impact of biological parameters on the *assortativity bias*.** Univariate sensitivity analysis of the odds ratios of infection between smokers and non-smokers varying: A) Probability of transmission per partnership, B) Duration of infection, and C) Probability of developing natural immunity after clearance of infection.

3.5.4 *Assortativity bias* assuming a true biological effect of smoking

Figure 3.4 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).

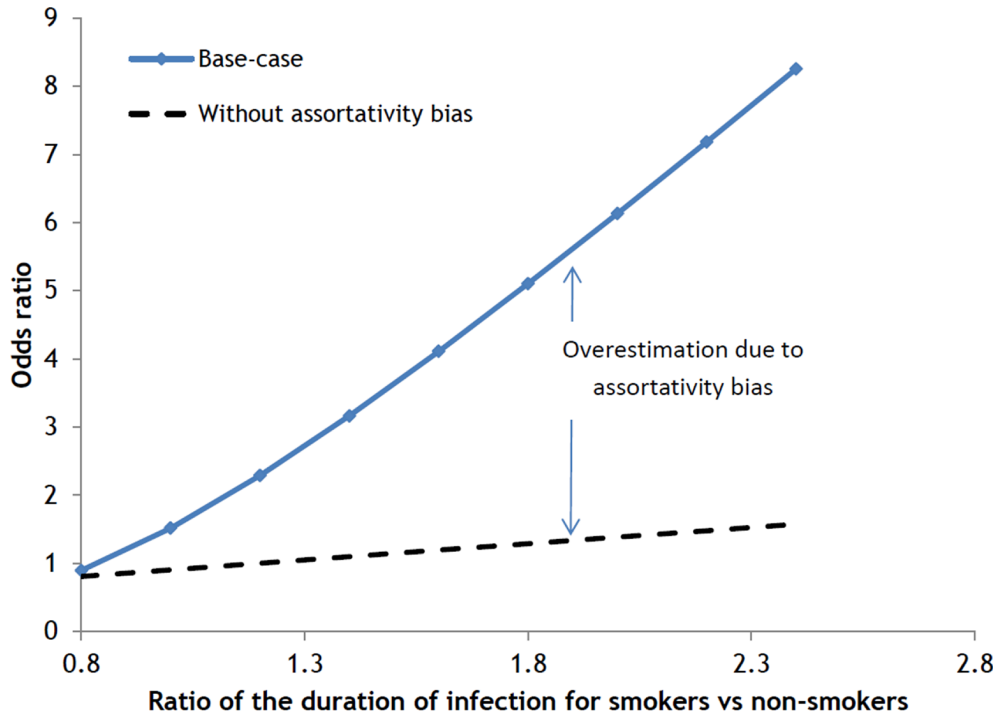


Figure 3.4 – **Impact of a biological effect of smoking on the duration of infection.** Univariate sensitivity analysis of the odds ratio (OR) of prevalence between smokers and non-smokers varying: Ratio of smokers' versus non-smokers' duration of infection. Two scenarios are shown both with base-case 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*.

3.6 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 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[235–238], 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 modeling 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%CI: 1.2, 1.7)[237], and most other studies have found ORs higher than 1.0[39;94;245;252]. 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 empirical-based studies for many other STIs[47;233;244]. We also showed that the size of the *assortativity bias* should be more important for STIs such as HIV, which have low R0.

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[253]. 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 behavior of participants (differences in the magnitude of C1-2).

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[41;250]. It is suggested that this is due to a biological cause (e.g., cervical ectopy makes young women vulnerable to STIs)[254]. Yet, sexual mixing is highly assortative with respect to age[247;254] (C1), and younger adults are more sexually active[105;254] (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 modeling to perfectly control a fictive population, allowing us 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) 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 Condition 2 that the association between sexual activity and smoking is monotonic, which seems to be the case[249]. 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[255;256], and sexual mixing is assortative by SES[256], 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 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[257;258], 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.

3.7 Conclusions

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.

Conclusion

3.8 Modelling multi-site HPV transmission and its impact on HPV vaccination effectiveness

3.8.1 Summary of results

The transmission-dynamic models of HPV infection used in analyses of population-level effectiveness and cost-effectiveness of HPV vaccination have included a single site of infection [169]. Hence, our first objective was to assess the impact of modelling HPV infection as a multi-site infection, with both a genital and an extragenital site, on model predictions of the population-level effectiveness of HPV vaccination against genital HPV infection (i.e., cervical HPV infection in women). We thus developed a uni-site and a multi-site HPV transmission model. When HPV is modelled as a multi-site infection, the natural history of infection becomes more complex and required more parameters to model. We needed to consider natural immunity that is site-specific (local immunity) and natural immunity that protects against infections at all sites (systemic immunity). Compared with the uni-site model, additional parameters were also needed to account for HPV transmission between different sites of infection (genital→extragenital, genital→genital, extragenital→extragenital, extragenital→genital). Using both the uni-site and multi-site models, we predicted the relative reduction in genital (cervico-vaginal) HPV16 prevalence post-vaccination among women. Because of the additional complexity in natural history of infection, the predictions made with the multi-site model were more uncertain than predictions made with the uni-site model. Assuming site-specific natural immunity, we found that the predicted reduction in HPV16 prevalence made with the multi-site model were always higher or equal to predictions made with the uni-site model. Assuming systemic natural immunity, predictions made with the multi-site model were always lower or equal to predictions made with the uni-site model.

These findings were further explained as stemming from three factors that result in higher predicted HPV vaccination effectiveness with the multi-site model: 1) higher proportion of genital infections caused by an extragenital infection (extragenital→genital transmission), 2) lower proportion of extragenital infections caused by a genital infection, and 3) higher proportion of susceptibles to extragenital infection.

Hence, we determined that the predicted reduction in genital HPV16 post-vaccination among women made with a uni-site model would be biased only if: A) natural immunity is site-specific and a substantial proportion of genital infections are caused by an extragenital infection, B) natural immunity is systemic and clearance of extragenital infection can result in acquisition of natural immunity.

We thus concluded that the current uni-site models used to assess the population-level effectiveness of HPV vaccination on genital diseases in women should be adequate. Indeed, the conditions under which the uni-site model would be biased are unlikely in light of the current understanding of HPV epidemiology. Although the current knowledge of HPV transmission and immunity is limited, the hypothesis most compatible with current data would be that natural immunity after cervical infection is systemic (i.e., serum antibodies), and that systemic natural immunity after extragenital infections is much less frequent. Indeed, seroconversion occurs much more frequently after clearance of cervical infections than after clearance of extragenital infections[27;29]. Furthermore, studies on HPV inter-site transmission conducted among couples have shown type-specific concordance of HPV infections between the oral and genital sites[201;259], but even greater concordance between the genital sites[201]. The oral-genital concordance in these cross-sectional studies could be due genital→extragenital transmission or extragenital→genital transmission. Hence, it is difficult to assess the potential of extragenital→genital transmission from the available evidence, but it appears as though the site mainly responsible for transmission of HPV to the cervico-vaginal site is the penis.

3.8.2 Strengths and limits

This study provides a conceptual understanding of the differences in the dynamics of transmission using a multi-site model compared to using a uni-site model. To do this, we considered three factors: 1) the proportion of individuals susceptibles to infection at each site of infection, 2) the magnitude of extragenital→genital transmission, and 3) the magnitude of genital→extragenital transmission. We were able to base our understanding of the impact of these factors on a theoretical analysis using simple homogeneous multi-site and uni-site models. Moreover, we illustrated the impact of these factors by calibrating the first multi-site model to HPV prevalence at the genital site. We have thus shown that these three factors explain the uncertainty in the predictions of the multi-site model.

There were two main limitations of the methodology presented in Chapter 1. Firstly, we did not use empirical data to inform directly the four probabilities of transmission in the multi-site model. Such data would include the probability of performing oral sex in a partnership, as well as the probability of oral HPV acquisition when performing oral sex on an individual infected at the genital site. Consequently, the uncertainty in the four probabilities of transmission was high, which translated into uncertainty in the predictions made with the multi-site model. However, we were still able to explain this uncertainty with three factors and thus evaluate the potential for bias in the uni-site model predictions. Secondly, we considered only two levels of sexual activity, we did not include stratification for age, and we only calibrated the transmission probabilities while keeping other parameters fixed.

Estimates of HPV vaccination population-level effectiveness are known to be sensitive to the probability of acquiring natural immunity upon clearance, and to the stratification of the model population in levels of sexual activity[169]. Therefore, our estimates of vaccination effectiveness may not be as accurate as those obtained with more heterogeneous models calibrated to extensive empirical data. However, we found that the relative reduction in HPV16 prevalence post-vaccination predicted by our uni-site model is comparable with the reduction predicted by other transmission-dynamic models of HPV infection[169].

3.8.3 Research in context and Future work

To our knowledge, two other published analyses have been done with a multi-site model[190;215]. However, this is the first analysis on the impact of multi-site transmission on predictions of HPV vaccination effectiveness. This has clear implications for predictions of the impact of HPV vaccination, which were estimated using only uni-site models[169]. Even though the current evidence does not support a significant bias in these predictions, we identified two important aspects of HPV natural history that need to be better understood before definitely excluding the possibility of significant bias: 1) the extent to which extragenital HPV infections can act as a reservoir for genital infections, 2) the existence and importance of local (cell-mediated) immunity as opposed to systemic (antibodies-mediated) immunity in the protection to subsequent HPV infections.

Recent sexual behavior surveys (e.g., NHANES and NSFG) have provided data on the practice of oral and anal sex. However, epidemiological data on the probability of HPV transmission through sexual acts such as oral sex, anal sex, vaginal sex or autoinoculation is currently lacking. Hence, we cannot expect to have well-informed transmission parameters in a multi-site model. The roles of autoinoculation, oral and anal sex, and sexual intercourse in the transmission between different sites need to be elucidated through longitudinal studies of HPV transmission. Until now, studies on HPV transmission have provided correlations between infections at different sites that are difficult to interpret meaningfully[88;90]. Multi-site transmission-dynamic modelling could be used to assist in the design as well as in the analysis of these studies in order to maximize the interpretability of the results. For example, a multi-site model can be used to simulate HPV transmission within the study population to determine the transmission parameters most compatible with the study results.

The need for better modelling of HPV as a site-specific infection will depend on the progression of vaccination coverage throughout the world. In countries with high vaccination coverage approaching the elimination threshold of genital HPV infection, the absolute bias from using a uni-site model would be small in any cases. In our study, a significant bias in the uni-site model predictions was observed up to a vaccination coverage of about 60%. In fact, at high vaccination coverage, the relative importance of herd immunity (indirect protection) in the population-level effectiveness of vaccination drops. Yet, the difference in predictions of effectiveness made with a uni-site model compared with a multi-site model stems entirely from the predicted herd immunity. The bias could still poten-

tially be significant in countries with lower coverage (<60%). Furthermore, if we aim to set optimal vaccination coverage targets in a post-elimination future, computing precise estimates of vaccination effectiveness for various coverages is critical. The bias from using a uni-site model could also matter in the decision of vaccinating boys, which depends, among other considerations, on the extent of herd effect from vaccinating girls on the burden of HPV-related diseases of the oral and anal sites.

Multi-site models may also have future applications in simulating the transmission of other site-specific STIs, such as HSV, gonorrhea, and chlamydia. In fact, it may not be possible to reproduce the transmission-dynamics and explain the spread of these STIs without including multi-site transmission. For instance, Hui et al.[215] have used a multi-site model to show that pharyngeal infections by gonorrhea have an important role in sustaining the transmission to the urethral site in a Men-who-have-Sex-with-Men. Hence, multi-site model may be needed to assess the impact of interventions on these site-specific STIs if herd immunity matters.

3.9 Impact of changes in sexual behavior on past and future trends of HPV infections and related cancers

3.9.1 Summary of results

The incidence of oropharyngeal cancer has increased more than 2-fold in US white men since the early 1980s[16], while the incidence of cervical cancer has decreased by around 50% since the 1970s in the US[218]. Our second objective was to assess the impact of changes in sexual behavior during the 20th century, including the so-called sexual revolution of the 1960s, on these trends of oropharyngeal and cervical cancers. We first documented the changes in sexual behavior observed in the NSFG, GSS and NHSLS data. We observed that the lifetime number of sexual partners has increased in both men and women. This increase is associated with a decrease in age at first sex and a lengthening of the pre-marital/pre-cohabitation period. We reproduced these changes in a dynamic model of HPV16 transmission and subsequent development of HPV16-related oropharyngeal and cervical cancers. We were thus able to produce simulations of the incidence of HPV16-related oropharyngeal and cervical cancers over the span of 130 years: from 1915 to 2045.

Our model predicted that, in the absence of cervical cancer screening, changes in lifetime number of sexual intercourse partners among women would have produced a 2-fold or more increase in cervical cancer between 1975 and 2015. Among men, our model predicted that the increase in the practice of oral sex has produced a 3-fold increase or more in HPV16-related oropharyngeal cancer. The predicted 3-fold increase in oropharyngeal cancer is comparable with the observed US trends in oropharyngeal cancer among white men. Our results also suggest that the increase in lifetime number of sexual intercourse partners among men is insufficient to explain the increase in oropharyngeal cancer. Finally, our model predicted that oropharyngeal cancer incidence could continue to increase up to 2045, which would affect unvaccinated cohorts born before 1990.

3.9.2 Strengths and limits

Our study is the first to use a transmission-dynamic model to assess the impact of trends in sexual behavior on incidence of HPV-related cancer. Previous analyses have used Age-Period-Cohort (APC) statistical modelling to understand trends in HPV infections and related cancers[217;218]. These types of models are based on statistical assumptions, such as the specification of a form for the relation between the sexual activity of a population and HPV-related outcomes[217]. However, such relation does not have a simple form or even a closed form, and these statistical assumptions cannot be easily related to the causal mechanisms. On the other hand, a mathematical model of transmission forces the relation between sexual activity and HPV-related outcomes to be based on causal mechanisms (i.e., transmission occurring during sexual partnerships from one infected to one susceptible), and the assumptions of the model can be related to these mechanisms (e.g., assumption of random sexual mixing). This extra structure prevents some of the misspecification that could occur with non-mechanistic statistical modelling.

Another strength of our approach is to have used data on cohabitation history to reproduce the sexual activity of older birth cohorts. Using data on the lifetime number of sexual partners alone is not enough, since such data do not provide information on the timing of sexual partnerships. However, since sexual life history and cohabitation history are strongly inter-related, we were able to use data on cohabitation history to estimate the timing of the sexual partnerships.

The important limits of this study concern the reproduction of sexual activity of the US population in the 20^{th} century. First, there is no longitudinal national survey on sexual activity, which means that sexual activity needed to be reconstructed through series of cross-sectional surveys from 1982 to 2015. Doing this required assumptions. Mainly, we had to assume that changes in sexual activity across birth cohorts could be completely captured by changes in cohabitation history (i.e., dates and lengths of cohabitations), in lifetime number of sexual partners, and in age at first sex. In other words, the sexual life history profiles of two women born at different dates, but with the same lifetime number of partners, the same cohabitation history and the same age at first sex is on average the same in the model population. If this hypothesis is wrong, there could be inaccuracies in the timing of sexual partners across the lifespan. However, the lifetime number of partners, the age at initiation, and the cohabitations would still be accurately reproduced. It seems unlikely that only a change of timing of sexual partners could explain our observations, such as the great discrepancy between our simulations of oropharyngeal cancer in scenario uni-site without oral sex and the empirical data (60%-80% increase vs 390%).

Second, the chance of reporting bias in the sexual behavior surveys is high. Under-reporting of sexual partners in women and over-reporting in men has been documented[113;117]and could affect differentially younger and older birth cohorts. We accounted for this potential bias by considering three different scenarios of under- and over-reporting in surveys. We have thus confirmed that a reporting

bias in men and women was the scenario providing the best fit for both oropharyngeal and cervical cancers. Third, because the historical data on the practice of oral sex are sparse, we cannot be certain that we reproduce with accuracy the practice of oral sex in the last century. In fact, we used a birth cohort effect to model the changes in the practice of oral sex, but we cannot exclude that a period effect could have also worked.

Finally, because of the age-limit of 44 years old in the model, we do not reproduce changes in sexual behavior that could have occurred in older age groups. This is important in light of the second and dominant peak in oral HPV prevalence observed between 50 and 60 years of age in NHANES[47]. If this second peak of oral HPV prevalence is due to acquisition of new infections at older age, then our simulations could be missing infections relevant for cancer. However, this hypothesis does not fit with the sexual behavior reported by aged men in surveys with no age limit (e.g., GSS). An alternative explanation for the second peak is that infection can go latent and reactivate when the immune system weaken at older ages.

3.9.3 Research in context and Future work

The rise in oropharyngeal cancer incidence has sparked interest in the practice of oral sex. Yet, it has been difficult to directly attribute oral HPV infection to oral sex, in part because of the correlation between the practice of oral sex and of sexual intercourse[47]. Some rare studies have managed to adjust for the number of vaginal sex partners[50;51] and show an association between the number of oral sex partners and oral HPV infection. Other studies have circumvented the problem of adjusting using different methods. For instance, one study showed that "barrier use during oral sex" among individuals practicing oral sex is associated with decreased oral HPV prevalence[86]. Another study made a comparison between ethnicities and birth cohorts with different sexual behaviors[260]. For instance, black Americans have higher lifetime number of vaginal sex partners than whites, but also have a lower number of partners on whom cunnilingus was performed. Furthermore, black Americans have lower oral HPV16 prevalence than whites, which suggests an excess risk of oral HPV from cunnilingus in white men[260]. In our work, we also add to this growing body of evidence through a different method: we show that sexual intercourse alone cannot explain changes in incidence of oropharyngeal cancer in white men, and that accounting for the practice of oral sex can produce the sharp increase observed in empirical data.

Our study is the first to account for changes in sexual behavior and provide evidence that without cervical cancer screening, there would have been a sharp increase in cervical cancer. The impact of screening on the incidence of cervical cancer was assessed at least in three other analyses[218;261;262]. The latest analysis[261] did not account for changes in sexual behavior, and so the estimated number of prevented cases due to screening is likely underestimated. In one other analysis in which an APC model was used[218], a birth cohort effect on cervical cancer incidence was masked by the period effect produced by the increase in screening coverage. The authors hypothesized that the birth cohort effect

could be due in part to changes in sexual behavior.

HPV transmission-dynamic models incorporating changes in sexual behavior have future applications for the assessment of the impact of HPV vaccination on HPV-related diseases[263]. Indeed, post-vaccination incidence of HPV-related diseases such as anogenital warts will depend not only on the parameters of the vaccination program (e.g., coverage, vaccine efficacy, etc.), but also on the underlying trends in incidence due to changes in sexual behavior. For example, we have shown that oropharyngeal cancer incidence is expected to increase in the next three decades in the absence of intervention. Hence, this trend of increasing incidence has to be teased out from the future incidence of oropharyngeal cancer to isolate the true impact of vaccination. Otherwise, we expect to underestimate the population-level effectiveness of vaccination. As we did in our study, simulations from transmission-dynamic models can provide the counterfactual of oropharyngeal cancer incidence in the absence of vaccination that is needed to assess the unbiased impact of vaccination.

To improve upon this work, we could include in the HPV transmission model each of the other three main ethnicity categories used in US surveys: Black, Hispanic, and Other. Since the different ethnicities have different sexual behavior and HPV-related outcomes, we can learn and validate our understanding of the relation between sexual behavior and HPV-related outcomes by including these populations in our model. For instance, we could assess, through simulations of HPV16 transmission, whether the lower frequency of oral sex practice among blacks or Hispanics can explain their lower prevalence of oral HPV16[260]. Furthermore, sexual mixing between whites and other ethnicities have increased throughout the past decades (these results were not shown), and this could also have played a role in the trends of HPV-related diseases among whites.

Another future avenue is to include in our HPV transmission model stratifications for the co-factors of progression to HPV-related cancers and for the risk factors of HPV infection. First, the prevalence or incidence of many of these factors (e.g., smoking, alcohol, contraceptive use, tonsillectomy) has changed during the past decades[136;264–266]. Therefore, the cumulative effect of these past changes on trends of HPV-related cancers may be significant and non-linear due to HPV being infectious. However, for many of these factors, the causal effect on progression to cancer or on acquisition/duration of HPV infection is difficult to establish. For example, factors such as smoking and alcohol have been found to be associated with oral HPV and are potential co-factors of progression from HPV infection to cancer[47;52;133;135]. Yet, sexual behavior and sexual mixing are also associated with alcohol and smoking[235;237;238;267], which can produce confounding when assessing the effect of these factors on HPV acquisition (see Chapter 3). This type of bias has been investigated and could be corrected using simulations from transmission-dynamic models integrating data on sexual behavior, sexual mixing, and stratification of the risk factor[99 268]. Thus, there is a future application for transmission-dynamic models in understanding the role of risk factors on HPV infection and progression to cancer, as well as the role of these factors in HPV-related cancer trends.

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

3.10.1 Summary of results

In epidemiological studies of STIs risk factors, adjustment is often only made for the individual-level characteristics of the study's subjects. Our final objective was to determine the conditions under which assortative mixing with respect to a potential risk factor could cause a bias persisting after adjustment for individual-level characteristics when estimating the causal effect of the risk factor on a STI. We used as a case study the association between HPV and smoking, which has been observed in many studies and interpreted as evidences of a biological effect of smoking on susceptibility to HPV or duration of infection through the immunosuppressive effects of tobacco[47]. We found two conditions that may lead to the *assortativity bias*: sexual mixing is assortative by smoking status (Condition 1) and smoking status is associated with sexual activity (Condition 2). We thus simulated the conduct of a cross-sectional study on the smoking - HPV association in which the Conditions 1-2 are present in the study population. We estimated the odds ratio (OR) of HPV prevalence comparing smokers with non-smokers. In our simulated study, even if smoking is not a biological cause of HPV infection, we observed an OR of 1.51 after perfectly adjusting for the subjects' individual-level characteristics. The deviation of the OR from the null value (1.00) was due to the *assortativity bias*. Assuming smoking had a biological effect on the probability of transmission or the duration of infection, then the OR adjusted for the subjects' individual-level characteristics still overestimated the true biological effect of smoking. We also observed that the magnitude of the bias increases as the degree of assortativity increases and as the strength of the association between sexual activity and smoking status increases. Finally, the magnitude of the bias, as measured by the OR, would be greater for STIs with lower R_0 such as HIV.

3.10.2 Strengths and limits

In this study, we were able to conceptualize the assortativity bias as arising under two conditions. Because of this, we are able to generalize our work to many other associations than the one between smoking and HPV. Indeed, Conditions 1 and 2 for the occurrence of the *assortativity bias* can be generalized by replacing smoking status by any risk factor and sexual activity by any cause of infection. The *assortativity bias* can also occur for STIs other than HPV. For example, young age is associated with cervical HPV independently of the number of lifetime partners[25]. Yet, sexual mixing is assortative for age[254], and young men have higher level of sexual activity than older men[254]. Thus, the two conditions for the *assortativity bias* are satisfied for the association between age and HPV infection. Likewise, sexual mixing is assortative with respect to ethnicity and black Americans have higher number of sexual partners than other ethnicities. Hence, black women have been observed to have higher rates of STIs than white women independently of their own number of sexual partners[19].

We were able to quantify the impact of this bias by using mathematical modelling and empirical data on smoking and sexual behavior to simulate a fictive study. There are two main strengths to this approach: 1) we circumvented the logistical and financial hurdles of conducting a study in which the HPV infection and smoking status of both the study subjects and subjects' sexual partners would be needed, 2) we can control all the parameters of the fictive study and population, which allows us to assess the impact of the bias in different contexts (e.g., high or low assortativity between smokers).

There are three main limits to this study. First, the transmission-dynamic model we used included only two levels of sexual activity. With additional levels of sexual activity in the model, we would need to specify in Condition 2 that sexual activity is monotonically associated with smoking status, which seems to be the case[249]. In other words, the proportion of smokers should grow as the level of sexual activity increases. Second, because we don't have empirical data to inform sexual mixing, mixing by level of sexual activity and smoking status was modelled using only two parameters: the degrees of assortativity for smoking and for the level of sexual activity. These two degrees of assortativity are insufficient to parameterize sexual mixing without further assumptions. For example, we implicitly assumed that assortative mixing by smoking status is independent of assortative mixing by level of sexual activity. We tested the impact of this assumption by including a parameter of correlation between the two types of assortativity (results not shown). Although the numerical value of the OR (comparing smokers with non-smokers) changed depending on the correlation, our qualitative conclusions were robust to this simplification. Finally, the fictive study and population were stratified for only two characteristics: smoking status and level of sexual activity. In epidemiological studies, multivariate adjustment for characteristics such as age, SES, and ethnicity is standard. Because the degree of assortativity by smoking status may decrease after stratification for these covariates, it is possible that the *assortativity bias* is partially adjusted for in epidemiological studies. For example, sexual mixing is assortative by age[254] and age is associated with smoking status[269]. Thus, the degree of assortativity with respect to smoking status may decrease after stratification for age.

3.10.3 Research in context and Future work

The traditional multivariate regression (logistic or cox) analysis of STI incidence or prevalence can be useful as a first step to identify important variables. However, such analysis is rarely adequate for causal inference on the variables because exposure to infection can vary across individuals[168]. In fact, the *assortativity bias* arises from a lack of control of exposure to infection. Indeed, in our simulations, the *assortativity bias* is due to smokers' higher probability of having a partner that smokes, which results in smokers having higher chance of being exposed to an infected partner. Yet, as in the case of HPV and smoking, the traditional multivariate analysis without control for exposure to infection is often both the starting and ending point of the assessment of risk factors[270;271]. This is because controlling for exposure to infection is logistically impossible in all but a few cases. Controlling for exposure can be done by either restricting the study population to subjects known to be exposed to an

infected partner[258], or by directly assessing the infection status of the subjects' sexual partners. When such designs are not feasible, we need to be able to better control for the probability of exposure to infection by using network-level variables.

In our simulated cross-sectional study on the HPV - smoking association, it is possible to control for the difference in probability of exposure to HPV infection between smokers and non-smokers. In fact, this difference in exposure to HPV stems entirely from smokers' sexual partners having a higher chance of being smokers and being highly sexually active compared to non-smokers' partners. Hence, if we could adjust for the smoking status of the subject's partner (as well as the subject's level of sexual activity), the imbalance in probability of exposure to infection between smokers and non-smokers would disappear. However, this strategy of adjustment may not work for more complicated sexual networks. Better knowledge of sexual network is needed to identify potential adjustment strategies. For instance, identifying the determinants of assortativity by level of sexual activity is crucial to be able to control for the partners' level of sexual activity without directly assessing it. Furthermore, there may be a use for ecological variables, such as the average socioeconomic status in the neighborhood, as a way to characterize the sexual network of subjects. To test which adjustment strategy is optimal for observational studies, smaller epidemiological studies of couples with assessment of infection status in both partners could be used[88;201;267]. With the progressive acquisition of data on sexual network, adjustment strategies can also be tested by simulating the conduct of a study with transmission dynamic models, just as we have done here.

3.10.4 Conclusion of the thesis

Traditionally, transmission-dynamic models have been used to predict the population-level effectiveness and cost-effectiveness of infectious disease prevention strategies. In this thesis, these types of models were rather used to understand epidemiological phenomena related to HPV: 1) impact of multi-site HPV transmission on the effectiveness of HPV vaccination, 2) impact of past trends in sexual behaviour on past and future incidence of HPV-related cancers, and 3) potential magnitude of *assortativity bias* in epidemiological studies examining the relationship between HPV infection and smoking. To do this, we developed three transmission-dynamic models that reproduced complex causal systems underlying HPV epidemiology, which enabled us to interpret empirical data without relying on the assumption of independence of outcomes (SUTVA[192]). By using models in such a way, we have shown that 1) models including only genital HPV transmission are likely adequate for predictions of the population-level impact of vaccination against genital HPV infection and related diseases, 2) changes in sexual behaviour over the past century could explain the rise in HPV-related oropharyngeal cancers, and 3) assortative mixing pattern may cause bias in the assessment of the causal role of smoking on HPV acquisition or duration of infection. Building on this work, HPV transmission-dynamic models can continue to be used as framework to interpret epidemiological data, and eventually be integrated in the design and analysis of epidemiological studies.

Bibliography

3.11 Bibliography

[1] J. M. Walboomers, M. V. Jacobs, M. M. Manos, F. X. Bosch, J. A. Kummer, K. V. Shah, P. J. Snijders, J. Peto, C. J. Meijer, and N. Munoz. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*, 189(1):12–9, 1999.

[2] H. De Vuyst, G. M. Clifford, M. C. Nascimento, M. M. Madeleine, and S. Franceschi. Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: a meta-analysis. *Int J Cancer*, 124(7):1626–36, 2009.

[3] A. R. Kreimer, G. M. Clifford, P. Boyle, and S. Franceschi. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev*, 14(2):467–75, 2005.

[4] C. J. Lacey, C. M. Lowndes, and K. V. Shah. Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease. *Vaccine*, 24 Suppl 3:S3/35–41, 2006.

[5] T. Goodman. Update on HPV vaccine introduction and programmatic perspectives, October 24 2018.

[6] M. Drolet, E. Benard, N. Perez, and M. Brisson. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. *Lancet*, 2019.

[7] D. Forman, C. de Martel, C. J. Lacey, I. Soerjomataram, J. Lortet-Tieulent, L. Bruni, J. Vignat, J. Ferlay, F. Bray, M. Plummer, and S. Franceschi. Global burden of human papillomavirus and related diseases. *Vaccine*, 30 Suppl 5:F12–23, 2012.

[8] M. Brisson, N. Van de Velde, and M. C. Boily. Economic evaluation of human papillomavirus vaccination in developed countries. *Public Health Genomics*, 12(5-6):343–51, 2009.

[9] S. O. Aral, K. A. Fenton, and K. K. Holmes. Sexually transmitted diseases in the USA: temporal trends. *Sex Transm Infect*, 83(4):257–66, 2007.

[10] G. Liu, S. Hariri, H. Bradley, S. L. Gottlieb, J. S. Leichliter, and L. E. Markowitz. Trends and patterns of sexual behaviors among adolescents and adults aged 14 to 59 years, United States. *Sex Transm Dis*, 42(1):20–6, 2015.

[11] J. H. Gagnon and W. Simon. The sexual scripting of oral genital contacts. *Arch Sex Behav*, 16(1):1–25, 1987.

[12] M. Veluire and D. Brasnu. [evolution of sexual behaviour in France and emergence of new head and neck cancers]. *Bull Cancer*, 98(10):1185–92, 2011.

[13] C. Fakhry and M. L. Gillison. Clinical implications of human papillomavirus in head and neck cancers. *J Clin Oncol*, 24(17):2606–11, 2006.

[14] A. K. Chaturvedi, E. A. Engels, W. F. Anderson, and M. L. Gillison. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. *J Clin Oncol*, 26(4):612–9, 2008.

[15] Wilkins K. Ellison LF. Canadian trends in cancer prevalence. *Statistics Canada, Catalogue no 82-003-XPE o Health Reports*, page 23, 2012.

[16] A. K. Chaturvedi, E. A. Engels, R. M. Pfeiffer, B. Y. Hernandez, W. Xiao, E. Kim, B. Jiang, M. T. Goodman, M. Sibug-Saber, W. Cozen, L. Liu, C. F. Lynch, N. Wentzensen, R. C. Jordan, S. Altekruse, W. F. Anderson, P. S. Rosenberg, and M. L. Gillison. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol*, 29(32):4294–301, 2011.

[17] J. C. Abma, G. M. Martinez, and C. E. Copen. Teenagers in the united states: sexual activity, contraceptive use, and childbearing, national survey of family growth 2006-2008. *Vital Health Stat 23*, (30):1–47, 2010.

[18] J. M. Twenge, R. A. Sherman, and B. E. Wells. Changes in American Adults' Sexual Behavior and Attitudes, 1972-2012. *Arch Sex Behav*, 44(8):2273–85, 2015.

[19] E. O. Laumann and Y. Youm. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. *Sex Transm Dis*, 26(5):250–61, 1999.

[20] A. Asiaf, S. T. Ahmad, S. O. Mohammad, and M. A. Zargar. Review of the current knowledge on the epidemiology, pathogenesis, and prevention of human papillomavirus infection. *Eur J Cancer Prev*, 23(3):206–24, 2014.

[21] E. M. de Villiers, C. Fauquet, T. R. Broker, H. U. Bernard, and H. zur Hausen. Classification of papillomaviruses. *Virology*, 324(1):17–27, 2004.

[22] N. Munoz, F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders, and C. J. Meijer. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*, 348(6):518–27, 2003.

[23] D. M. Parkin and F. Bray. Chapter 2: The burden of HPV-related cancers. *Vaccine*, 24 Suppl 3:S3/11–25, 2006.

[24] C. B. Woodman, S. I. Collins, and L. S. Young. The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer*, 7(1):11–22, 2007.

[25] E. F. Dunne, E. R. Unger, M. Sternberg, G. McQuillan, D. C. Swan, S. S. Patel, and L. E. Markowitz. Prevalence of HPV infection among females in the United States. *JAMA*, 297(8):813–9, 2007.

[26] A. A. Deshmukh, R. J. Tanner, M. C. Luetke, Y. R. Hong, K. Sonawane Deshmukh, and 3rd Mainous, A. G. Prevalence and Risk of Penile Human Papillomavirus Infection: Evidence From The National Health and Nutrition Examination Survey 2013-2014. *Clin Infect Dis*, 64(10):1360–1366, 2017.

[27] J. J. Carter, L. A. Koutsky, J. P. Hughes, S. K. Lee, J. Kuypers, N. Kiviat, and D. A. Galloway. Comparison of human papillomavirus types 16, 18, and 6 capsid antibody responses following incident infection. *J Infect Dis*, 181(6):1911–9, 2000.

[28] Z. R. Edelstein, J. J. Carter, R. Garg, R. L. Winer, Q. Feng, D. A. Galloway, and L. A. Koutsky. Serum antibody response following genital alpha9 human papillomavirus infection in young men. *J Infect Dis*, 204(2):209–16, 2011.

[29] A. R. Giuliano, R. Viscidi, B. N. Torres, D. J. Ingles, S. L. Sudenga, L. L. Villa, M. L. Baggio, M. Abrahamsen, M. Quiterio, J. Salmeron, and E. Lazcano-Ponce. Seroconversion Following Anal and Genital HPV Infection in Men: The HIM Study. *Papillomavirus Res*, 1:109–115, 2015.

[30] G. Y. Ho, R. Bierman, L. Beardsley, C. J. Chang, and R. D. Burk. Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med*, 338(7):423–8, 1998.

[31] C. B. Woodman, S. Collins, H. Winter, A. Bailey, J. Ellis, P. Prior, M. Yates, T. P. Rollason, and L. S. Young. Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. *Lancet*, 357(9271):1831–6, 2001.

[32] A. R. Giuliano, J. H. Lee, W. Fulp, L. L. Villa, E. Lazcano, M. R. Papenfuss, M. Abrahamsen, J. Salmeron, G. M. Anic, D. E. Rollison, and D. Smith. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. *Lancet*, 377(9769):932–40, 2011.

[33] Jr. Moreira, E. D., A. R. Giuliano, J. Palefsky, C. A. Flores, S. Goldstone, D. Ferris, R. J. Hillman, H. Moi, M. H. Stoler, B. Marshall, S. Vuocolo, D. Guris, and R. M. Haupt. Incidence, clearance, and disease progression of genital human papillomavirus infection in heterosexual men. *J Infect Dis*, 210(2):192–9, 2014.

[34] A. B. Moscicki, N. Hills, S. Shiboski, K. Powell, N. Jay, E. Hanson, S. Miller, L. Clayton, S. Farhat, J. Broering, T. Darragh, and J. Palefsky. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. *JAMA*, 285(23):2995–3002, 2001.

[35] M. Steinau, S. Hariri, M. L. Gillison, T. R. Broutian, E. F. Dunne, Z. Y. Tong, L. E. Markowitz, and E. R. Unger. Prevalence of cervical and oral human papillomavirus infections among US women. *J Infect Dis*, 209(11):1739–43, 2014.

[36] P. Oakeshott, A. Aghaizu, F. Reid, R. Howell-Jones, P. E. Hay, S. T. Sadiq, C. J. Lacey, S. Beddows, and K. Soldan. Frequency and risk factors for prevalent, incident, and persistent genital carcinogenic human papillomavirus infection in sexually active women: community based cohort study. *BMJ*, 344:e4168, 2012.

[37] E. Roset Bahmanyar, J. Paavonen, P. Naud, J. Salmeron, S. N. Chow, D. Apter, H. Kitchener, X. Castellsague, J. C. Teixeira, S. R. Skinner, U. Jaisamrarn, G. A. Limson, S. M. Garland, A. Szarewski, B. Romanowski, F. Aoki, T. F. Schwarz, W. A. Poppe, N. S. De Carvalho, D. M. Harper, F. X. Bosch, A. Raillard, D. Descamps, F. Struyf, M. Lehtinen, and G. Dubin. Prevalence and risk factors for cervical HPV infection and abnormalities in young adult women at enrolment in the multinational PATRICIA trial. *Gynecol Oncol*, 127(3):440–50, 2012.

[38] C. Remschmidt, A. M. Kaufmann, I. Hagemann, E. Vartazarova, O. Wichmann, and Y. Delere. Risk factors for cervical human papillomavirus infection and high-grade intraepithelial lesion in women aged 20 to 31 years in Germany. *Int J Gynecol Cancer*, 23(3):519–26, 2013.

[39] C. Ley, H. M. Bauer, A. Reingold, M. H. Schiffman, J. C. Chambers, C. J. Tashiro, and M. M. Manos. Determinants of genital human papillomavirus infection in young women. *J Natl Cancer Inst*, 83(14):997–1003, 1991.

[40] A. R. Giuliano, M. Papenfuss, A. Schneider, M. Nour, and K. Hatch. Risk factors for high-risk type human papillomavirus infection among Mexican-American women. *Cancer Epidemiol Biomarkers Prev*, 8(7):615–20, 1999.

[41] C. M. Wheeler, W. C. Hunt, J. Cuzick, E. Langsfeld, A. Pearse, G. D. Montoya, M. Robertson, C. A. Shearman, and P. E. Castle. A population-based study of human papillomavirus genotype prevalence in the United States: baseline measures prior to mass human papillomavirus vaccination. *Int J Cancer*, 132(1):198–207, 2013.

[42] A. G. Nyitray, R. J. da Silva, M. L. Baggio, B. Lu, D. Smith, M. Abrahamsen, M. Papenfuss, M. Quiterio, L. L. Villa, and A. R. Giuliano. The prevalence of genital HPV and factors associated with oncogenic HPV among men having sex with men and men having sex with women and men: the HIM study. *Sex Transm Dis*, 38(10):932–40, 2011.

[43] E. I. Svare, S. K. Kjaer, A. M. Worm, A. Osterlind, C. J. Meijer, and A. J. van den Brule. Risk factors for genital HPV DNA in men resemble those found in women: a study of male attendees at a Danish STD clinic. *Sex Transm Infect*, 78(3):215–8, 2002.

[44] V. Colon-Lopez, A. P. Ortiz, L. Del Toro-Mejias, M. Clatts, G. Duran-Guzman, N. Perez, M. DaCosta, and J. Palefsky. Prevalence and Correlates of Penile HPV Infection in a Clinic-Based Sample of Hispanic Males. *P R Health Sci J*, 34(3):128–34, 2015.

[45] F. X. Wei, M. Guo, X. J. Ma, Y. Huang, Y. Zheng, L. Wang, Y. Sun, S. J. Zhuang, K. Yin, Y. Y. Su, S. J. Huang, M. Q. Li, T. Wu, and J. Zhang. [The impact of male circumcision on the natural history of genital HPV infection: a prospective cohort study]. *Zhonghua Yu Fang Yi Xue Za Zhi*, 52(5):486–492, 2018.

[46] T. B. Olesen, J. Mwaiselage, T. Iftner, C. Kahesa, V. Rasch, K. Frederiksen, C. Munk, and S. K. Kjaer. Risk factors for genital human papillomavirus among men in Tanzania. *J Med Virol*, 89(2):345–351, 2017.

[47] M. L. Gillison, T. Broutian, R. K. Pickard, Z. Y. Tong, W. Xiao, L. Kahle, B. I. Graubard, and A. K. Chaturvedi. Prevalence of oral HPV infection in the United States, 2009-2010. *JAMA*, 307(7):693–703, 2012.

[48] G. D'Souza, A. Wentz, N. Kluz, Y. Zhang, E. Sugar, R. M. Youngfellow, Y. Guo, W. Xiao, and M. L. Gillison. Sex Differences in Risk Factors and Natural History of Oral Human Papillomavirus Infection. *J Infect Dis*, 213(12):1893–6, 2016.

[49] A. R. Kreimer, C. M. Pierce Campbell, H. Y. Lin, W. Fulp, M. R. Papenfuss, M. Abrahamsen, A. Hildesheim, L. L. Villa, J. J. Salmeron, E. Lazcano-Ponce, and A. R. Giuliano. Incidence and clearance of oral human papillomavirus infection in men: the HIM cohort study. *Lancet*, 382(9895):877–87, 2013.

[50] R. K. Pickard, W. Xiao, T. R. Broutian, X. He, and M. L. Gillison. The prevalence and incidence of oral human papillomavirus infection among young men and women, aged 18-30 years. *Sex Transm Dis*, 39(7):559–66, 2012.

[51] G. D'Souza, Y. Agrawal, J. Halpern, S. Bodison, and M. L. Gillison. Oral sexual behaviors associated with prevalent oral human papillomavirus infection. *J Infect Dis*, 199(9):1263–9, 2009.

[52] A. R. Kreimer, A. Villa, A. G. Nyitray, M. Abrahamsen, M. Papenfuss, D. Smith, A. Hildesheim, L. L. Villa, E. Lazcano-Ponce, and A. R. Giuliano. The epidemiology of oral HPV infection among a multinational sample of healthy men. *Cancer Epidemiol Biomarkers Prev*, 20(1):172–82, 2011.

[53] C. H. Chung, A. Bagheri, and G. D'Souza. Epidemiology of oral human papillomavirus infection. *Oral Oncol*, 50(5):364–9, 2014.

[54] B. Y. Hernandez, K. McDuffie, X. Zhu, L. R. Wilkens, J. Killeen, B. Kessel, M. T. Wakabayashi, C. C. Bertram, D. Easa, L. Ning, J. Boyd, C. Sunoo, L. Kamemoto, and M. T. Goodman. Anal human papillomavirus infection in women and its relationship with cervical infection. *Cancer Epidemiol Biomarkers Prev*, 14(11 Pt 1):2550–6, 2005.

[55] F. A. Castro, W. Quint, P. Gonzalez, H. A. Katki, R. Herrero, L. J. van Doorn, M. Schiffman, L. Struijk, A. C. Rodriguez, C. DelVecchio, D. R. Lowy, C. Porras, S. Jimenez, J. Schiller, D. Solomon, S. Wacholder, A. Hildesheim, and A. R. Kreimer. Prevalence of and risk factors for anal human papillomavirus infection among young healthy women in Costa Rica. *J Infect Dis*, 206(7):1103–10, 2012.

[56] A. G. Nyitray, R. J. Carvalho da Silva, M. L. Baggio, B. Lu, D. Smith, M. Abrahamsen, M. Papenfuss, L. L. Villa, E. Lazcano-Ponce, and A. R. Giuliano. Age-specific prevalence of and risk factors for anal human papillomavirus (HPV) among men who have sex with women and men who have sex with men: the HPV in men (HIM) study. *J Infect Dis*, 203(1):49–57, 2011.

[57] Y. B. Shvetsov, B. Y. Hernandez, K. McDuffie, L. R. Wilkens, X. Zhu, L. Ning, J. Killeen, L. Kamemoto, and M. T. Goodman. Duration and clearance of anal human papillomavirus (HPV) infection among women: the hawaii HPV cohort study. *Clin Infect Dis*, 48(5):536–46, 2009.

[58] M. T. Goodman, Y. B. Shvetsov, K. McDuffie, L. R. Wilkens, X. Zhu, P. J. Thompson, L. Ning, J. Killeen, L. Kamemoto, and B. Y. Hernandez. Sequential acquisition of human papillomavirus (HPV) infection of the anus and cervix: the Hawaii HPV Cohort study. *J Infect Dis*, 201(9):1331–9, 2010.

[59] T. Guler, D. Uygur, M. Uncu, E. Yayci, T. Atacag, K. Bas, M. Gunay, and C. Yakicier. Coexisting anal human papilloma virus infection in heterosexual women with cervical HPV infection. *Arch Gynecol Obstet*, 288(3):667–72, 2013.

[60] Jr. Simpson, S., P. Blomfield, A. Cornall, S. N. Tabrizi, L. Blizzard, and R. Turner. Front-to-back & dabbing wiping behaviour post-toilet associated with anal neoplasia & HR-HPV carriage in women with previous HPV-mediated gynaecological neoplasia. *Cancer Epidemiol*, 42:124–132, 2016.

[61] T. Tian, P. Mijiti, H. Bingxue, Z. Fadong, A. Ainiwaer, S. Guoyao, Z. Zhanlin, Y. Mahan, T. Xiaoqin, G. Zheng, and D. Jianghong. Prevalence and risk factors of anal human papillomavirus infection among HIV-negative men who have sex with men in Urumqi city of Xinjiang Uyghur Autonomous Region, China. *PLoS One*, 12(11):e0187928, 2017.

[62] M. G. Dona, M. F. Vescio, A. Latini, A. Giglio, D. Moretto, M. Frasca, M. Benevolo, F. Rollo, M. Colafigli, A. Cristaudo, and M. Giuliani. Anal human papillomavirus in HIV-uninfected men who have sex with men: incidence and clearance rates, duration of infection, and risk factors. *Clin Microbiol Infect*, 22(12):1004 e1–1004 e7, 2016.

[63] R. D. Cranston, A. D. Althouse, F. van Griensven, L. Janocko, M. E. Curlin, S. Chaikummao, W. Chonwattana, A. Siegel, T. H. Holtz, and I. McGowan. Prevalence of Anal Human Papillomavirus Vaccine Types in the Bangkok Men Who Have Sex With Men Cohort Study. *Sex Transm Dis*, 42(12):671–6, 2015.

[64] D. J. Wiley, X. Li, H. Hsu, E. C. Seaberg, R. D. Cranston, S. Young, G. D'Souza, O. Martinez-Maza, K. DeAzambuja, K. Chua, S. K. Hussain, and R. Detels. Factors affecting the prevalence of strongly and weakly carcinogenic and lower-risk human papillomaviruses in anal specimens in a cohort of men who have sex with men (msm). *PLoS One*, 8(11):e79492, 2013.

[65] F. van Aar, S. H. Mooij, M. A. van der Sande, A. G. Speksnijder, I. G. Stolte, C. J. Meijer, D. W. Verhagen, A. J. King, H. J. de Vries, and M. F. Schim van der Loeff. Anal and penile high-risk human papillomavirus prevalence in HIV-negative and HIV-infected MSM. *AIDS*, 27(18):2921–31, 2013.

[66] F. X. Bosch and S. de Sanjose. The epidemiology of human papillomavirus infection and cervical cancer. *Dis Markers*, 23(4):213–27, 2007.

[67] S. K. Kjaer, B. Chackerian, A. J. van den Brule, E. I. Svare, G. Paull, J. M. Walbomers, J. T. Schiller, J. E. Bock, M. E. Sherman, D. R. Lowy, and C. L. Meijer. High-risk human papillomavirus is sexually transmitted: evidence from a follow-up study of virgins starting sexual activity (intercourse). *Cancer Epidemiol Biomarkers Prev*, 10(2):101–6, 2001.

[68] J. S. Smith, A. Melendy, R. K. Rana, and J. M. Pimenta. Age-specific prevalence of infection with human papillomavirus in females: a global review. *J Adolesc Health*, 43(4 Suppl):S5–25, S25 e1–41, 2008.

[69] S. de Sanjose, M. Diaz, X. Castellsague, G. Clifford, L. Bruni, N. Munoz, and F. X. Bosch. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *Lancet Infect Dis*, 7(7):453–9, 2007.

[70] S. Franceschi and I. Baussano. Naturally acquired immunity against human papillomavirus (HPV): why it matters in the HPV vaccine era. *J Infect Dis*, 210(4):507–9, 2014.

[71] C. Velicer, X. Zhu, S. Vuocolo, K. L. Liaw, and A. Saah. Prevalence and incidence of HPV genital infection in women. *Sex Transm Dis*, 36(11):696–703, 2009.

[72] P. E. Gravitt and R. L. Winer. Natural History of HPV Infection across the Lifespan: Role of Viral Latency. *Viruses*, 9(10), 2017.

[73] S. L. Sudenga and S. Shrestha. Key considerations and current perspectives of epidemiological studies on human papillomavirus persistence, the intermediate phenotype to cervical cancer. *Int J Infect Dis*, 17(4):e216–20, 2013.

[74] P. E. Gravitt. The known unknowns of HPV natural history. *J Clin Invest*, 121(12):4593–9, 2011.

[75] E. F. Dunne, C. M. Nielson, K. M. Stone, L. E. Markowitz, and A. R. Giuliano. Prevalence of HPV infection among men: A systematic review of the literature. *J Infect Dis*, 194(8):1044–57, 2006.

[76] J. S. Smith, P. A. Gilbert, A. Melendy, R. K. Rana, and J. M. Pimenta. Age-specific prevalence of human papillomavirus infection in males: a global review. *J Adolesc Health*, 48(6):540–52, 2011.

[77] D. C. Beachler, L. A. Pinto, T. J. Kemp, A. G. Nyitray, A. Hildesheim, R. Viscidi, J. Schussler, A. R. Kreimer, and A. R. Giuliano. An Examination of HPV16 Natural Immunity in Men Who Have Sex with Men (MSM) in the HPV in Men (HIM) Study. *Cancer Epidemiol Biomarkers Prev*, 27(4):496–502, 2018.

[78] S. L. Ranjeva, E. B. Baskerville, V. Dukic, L. L. Villa, E. Lazcano-Ponce, A. R. Giuliano, G. Dwyer, and S. Cobey. Recurring infection with ecologically distinct HPV types can explain high prevalence and diversity. *Proc Natl Acad Sci U S A*, 114(51):13573–13578, 2017.

[79] A. R. Kreimer, A. J. Alberg, R. Viscidi, and M. L. Gillison. Gender differences in sexual biomarkers and behaviors associated with human papillomavirus-16, -18, and -33 seroprevalence. *Sex Transm Dis*, 31(4):247–56, 2004.

[80] S. Tam, S. Fu, L. Xu, K. J. Krause, D. R. Lairson, H. Miao, E. M. Sturgis, and K. R. Dahlstrom. The epidemiology of oral human papillomavirus infection in healthy populations: A systematic review and meta-analysis. *Oral Oncol*, 82:91–99, 2018.

[81] D. C. Beachler, K. A. Lang Kuhs, L. Struijk, J. Schussler, R. Herrero, C. Porras, A. Hildesheim, B. Cortes, J. Sampson, W. Quint, P. Gonzalez, and A. R. Kreimer. The Natural History of Oral Human Papillomavirus in Young Costa Rican Women. *Sex Transm Dis*, 44(7):442–449, 2017.

[82] D. C. Beachler, E. A. Sugar, J. B. Margolick, K. M. Weber, H. D. Strickler, D. J. Wiley, R. D. Cranston, R. D. Burk, H. Minkoff, S. Reddy, W. Xiao, Y. Guo, M. L. Gillison, and G. D'Souza.

Risk factors for acquisition and clearance of oral human papillomavirus infection among HIV-infected and HIV-uninfected adults. *Am J Epidemiol*, 181(1):40–53, 2015.

[83] C. Fakhry, M. L. Gillison, and G. D'Souza. Tobacco use and oral HPV-16 infection. *JAMA*, 312(14):1465–7, 2014.

[84] F. van Aar, S. H. Mooij, M. A. van der Sande, C. J. Meijer, A. J. King, D. W. Verhagen, T. Heijman, R. A. Coutinho, and M. F. Schim van der Loeff. Twelve-month incidence and clearance of oral HPV infection in HIV-negative and HIV-infected men who have sex with men: the H2M cohort study. *BMC Infect Dis*, 14:668, 2014.

[85] Z. R. Edelstein, S. M. Schwartz, S. Hawes, J. P. Hughes, Q. Feng, M. E. Stern, S. O'Reilly, S. K. Lee, L. Fu Xi, and L. A. Koutsky. Rates and determinants of oral human papillomavirus infection in young men. *Sex Transm Dis*, 39(11):860–7, 2012.

[86] A. Gupta, R. B. Perkins, G. Ortega, S. Feldman, and A. Villa. Barrier use during oro-genital sex and oral Human Papillomavirus prevalence: Analysis of NHANES 2009-2014. *Oral Dis*, 2018.

[87] S. H. Mooij, O. Landen, F. R. van der Klis, M. A. van der Sande, H. E. de Melker, M. Xiridou, A. van Eeden, T. Heijman, A. G. Speksnijder, P. J. Snijders, and M. F. Schim van der Loeff. HPV seroconversion following anal and penile HPV infection in HIV-negative and HIV-infected MSM. *Cancer Epidemiol Biomarkers Prev*, 23(11):2455–61, 2014.

[88] B. Y. Hernandez, L. R. Wilkens, X. Zhu, P. Thompson, K. McDuffie, Y. B. Shvetsov, L. E. Kamemoto, J. Killeen, L. Ning, and M. T. Goodman. Transmission of human papillomavirus in heterosexual couples. *Emerg Infect Dis*, 14(6):888–94, 2008.

[89] E. J. Ryndock and C. Meyers. A risk for non-sexual transmission of human papillomavirus? *Expert Rev Anti Infect Ther*, 12(10):1165–70, 2014.

[90] T. Malagon, K. Louvanto, M. Wissing, A. N. Burchell, P. P. Tellier, M. El-Zein, F. Coutlee, and E. L. Franco. Hand-to-genital and genital-to-genital transmission of human papillomaviruses between male and female sexual partners (HITCH): a prospective cohort study. *Lancet Infect Dis*, 19(3):317–326, 2019.

[91] R. L. Cook, E. L. Thompson, N. E. Kelso, J. Friary, J. Hosford, P. Barkley, V. J. Dodd, M. Abrahamsen, S. Ajinkya, P. D. Obesso, M. H. Rashid, and A. R. Giuliano. Sexual behaviors and other risk factors for oral human papillomavirus infections in young women. *Sex Transm Dis*, 41(8):486–92, 2014.

[92] N. Termine, L. Giovannelli, D. Matranga, M. P. Caleca, C. Bellavia, A. Perino, and G. Campisi. Oral human papillomavirus infection in women with cervical HPV infection: new data from an Italian cohort and a metanalysis of the literature. *Oral Oncol*, 47(4):244–50, 2011.

[93] L. A. Brinton, W. C. Reeves, M. M. Brenes, R. Herrero, E. Gaitan, F. Tenorio, R. C. de Britton, M. Garcia, and W. E. Rawls. The male factor in the etiology of cervical cancer among sexually monogamous women. *Int J Cancer*, 44(2):199–203, 1989.

[94] R. D. Burk, G. Y. Ho, L. Beardsley, M. Lempa, M. Peters, and R. Bierman. Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women. *J Infect Dis*, 174(4):679–89, 1996.

[95] S. K. Kjaer, E. M. de Villiers, C. Dahl, G. Engholm, J. E. Bock, B. F. Vestergaard, E. Lynge, and O. M. Jensen. Case-control study of risk factors for cervical neoplasia in Denmark. I: Role of the "male factor" in women with one lifetime sexual partner. *Int J Cancer*, 48(1):39–44, 1991.

[96] A. B. Moscicki, Y. Ma, J. Jonte, S. Miller-Benningfield, E. Hanson, J. Jay, C. Godwin de Medina, S. Farhat, L. Clayton, and S. Shiboski. The role of sexual behavior and human papillomavirus persistence in predicting repeated infections with new human papillomavirus types. *Cancer Epidemiol Biomarkers Prev*, 19(8):2055–65, 2010.

[97] M. C. Rousseau, J. S. Pereira, J. C. Prado, L. L. Villa, T. E. Rohan, and E. L. Franco. Cervical coinfection with human papillomavirus (HPV) types as a predictor of acquisition and persistence of HPV infection. *J Infect Dis*, 184(12):1508–17, 2001.

[98] F. Mendez, N. Munoz, H. Posso, M. Molano, V. Moreno, A. J. van den Brule, M. Ronderos, C. Meijer, and A. Munoz. Cervical coinfection with human papillomavirus (HPV) types and possible implications for the prevention of cervical cancer by HPV vaccines. *J Infect Dis*, 192(7):1158–65, 2005.

[99] T. Malagon, P. Lemieux-Mellouki, J. F. Laprise, and M. Brisson. Bias Due to Correlation Between Times-at-Risk for Infection in Epidemiologic Studies Measuring Biological Interactions Between Sexually Transmitted Infections: A Case Study Using Human Papillomavirus Type Interactions. *Am J Epidemiol*, 184(12):873–883, 2016.

[100] S. Strauss, P. Sastry, C. Sonnex, S. Edwards, and J. Gray. Contamination of environmental surfaces by genital human papillomaviruses. *Sex Transm Infect*, 78(2):135–8, 2002.

[101] S. H. Liu, R. M. Brotman, J. M. Zenilman, P. E. Gravitt, and D. A. Cummings. Menstrual cycle and detectable human papillomavirus in reproductive-age women: a time series study. *J Infect Dis*, 208(9):1404–15, 2013.

[102] M. C. Bleeker, C. J. Hogewoning, J. Berkhof, F. J. Voorhorst, A. T. Hesselink, P. M. van Diemen, A. J. van den Brule, P. J. Snijders, and C. J. Meijer. Concordance of specific human papillomavirus types in sex partners is more prevalent than would be expected by chance and is associated with increased viral loads. *Clin Infect Dis*, 41(5):612–20, 2005.

[103] C. F. Turner, R. D. Danella, and S. M. Rogers. Sexual behavior in the United States 1930-1990: trends and methodological problems. *Sex Transm Dis*, 22(3):173–90, 1995.

[104] D. Herbenick, M. Reece, V. Schick, S. A. Sanders, B. Dodge, and J. D. Fortenberry. Sexual behavior in the United States: results from a national probability sample of men and women ages 14-94. *J Sex Med*, 7 Suppl 5:255–65, 2010.

[105] A. Chandra, W. D. Mosher, C. Copen, and C. Sionean. 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*, (36):1–36, 2011.

[106] A. E. Biddlecom. Trends in sexual behaviours and infections among young people in the United States. *Sex Transm Infect*, 80 Suppl 2:ii74–9, 2004.

[107] J. S. Santelli, L. D. Lindberg, J. Abma, C. S. McNeely, and M. Resnick. Adolescent sexual behavior: estimates and trends from four nationally representative surveys. *Fam Plann Perspect*, 32(4):156–65, 194, 2000.

[108] D. K. Eaton, L. Kann, S. Kinchen, S. Shanklin, K. H. Flint, J. Hawkins, W. A. Harris, R. Lowry, T. McManus, D. Chyen, L. Whittle, C. Lim, and H. Wechsler. Youth risk behavior surveillance - United States, 2011. *MMWR Surveill Summ*, 61(4):1–162, 2012.

[109] D. J. Besharov and K. N. Gardiner. Trends in teen sexual behavior. *Child Youth Serv Rev*, 19(5-6):341–67, 1997.

[110] W. D. Mosher, A. Chandra, and J. Jones. Sexual behavior and selected health measures: men and women 15-44 years of age, United States, 2002. *Adv Data*, (362):1–55, 2005.

[111] L. Remez. Oral sex among adolescents: is it sex or is it abstinence? *Fam Plann Perspect*, 32(6):298–304, 2000.

[112] G. J. Gates and F. L. Sonenstein. Heterosexual genital sexual activity among adolescent males: 1988 and 1995. *Fam Plann Perspect*, 32(6):295–7, 304, 2000.

[113] T. D. Fisher. Sex of experimenter and social norm effects on reports of sexual behavior in young men and women. *Arch Sex Behav*, 36(1):89–100, 2007.

[114] D. A. Kreager and J. Staff. THE SEXUAL DOUBLE STANDARD AND ADOLESCENT PEER ACCEPTANCE. *Soc Psychol Q*, 72(2):143–164, 2009.

[115] T. W. Smith. Discrepancies between men and women in reporting number of sexual partners: a summary from four countries. *Soc Biol*, 39(3-4):203–11, 1992.

[116] K. R. Mitchell, C. H. Mercer, P. Prah, S. Clifton, C. Tanton, K. Wellings, and A. Copas. Why Do Men Report More Opposite-Sex Sexual Partners Than Women? Analysis of the Gender Discrepancy in a British National Probability Survey. *J Sex Res*, pages 1–8, 2018.

[117] M. G. Alexander and T. D. Fisher. Truth and consequences: using the bogus pipeline to examine sex differences in self-reported sexuality. *J Sex Res*, 40(1):27–35, 2003.

[118] Peter K. Jonason and Terri D. Fisher. The Power of Prestige: Why Young Men Report Having more Sex Partners than Young Women. *Sex Roles*, 60(3):151–159, 2009.

[119] D. D. Brewer, J. J. Potterat, S. B. Garrett, S. Q. Muth, Jr. Roberts, J. M., D. Kasprzyk, D. E. Montano, and W. W. Darrow. Prostitution and the sex discrepancy in reported number of sexual partners. *Proc Natl Acad Sci U S A*, 97(22):12385–8, 2000.

[120] A. B. Moscicki, M. Schiffman, A. Burchell, G. Albero, A. R. Giuliano, M. T. Goodman, S. K. Kjaer, and J. Palefsky. Updating the natural history of human papillomavirus and anogenital cancers. *Vaccine*, 30 Suppl 5:F24–33, 2012.

[121] J. Melnikow, J. Nuovo, A. R. Willan, B. K. Chan, and L. P. Howell. Natural history of cervical squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol*, 92(4 Pt 2):727–35, 1998.

[122] M. Plummer, R. Herrero, S. Franceschi, C. J. Meijer, P. Snijders, F. X. Bosch, S. de Sanjose, and N. Munoz. Smoking and cervical cancer: pooled analysis of the IARC multi-centric case–control study. *Cancer Causes Control*, 14(9):805–14, 2003.

[123] V. Moreno, F. X. Bosch, N. Munoz, C. J. Meijer, K. V. Shah, J. M. Walboomers, R. Herrero, and S. Franceschi. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. *Lancet*, 359(9312):1085–92, 2002.

[124] S. Collins, T. P. Rollason, L. S. Young, and C. B. Woodman. Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: a longitudinal study. *Eur J Cancer*, 46(2):405–11, 2010.

[125] A. B. Bos, M. Rebolj, J. D. Habbema, and M. van Ballegooijen. Nonattendance is still the main limitation for the effectiveness of screening for cervical cancer in the Netherlands. *Int J Cancer*, 119(10):2372–5, 2006.

[126] U. Jaisamrarn, X. Castellsague, S. M. Garland, P. Naud, J. Palmroth, M. R. Del Rosario-Raymundo, C. M. Wheeler, J. Salmeron, S. N. Chow, D. Apter, J. C. Teixeira, S. R. Skinner, J. Hedrick, A. Szarewski, B. Romanowski, F. Y. Aoki, T. F. Schwarz, W. A. Poppe, F. X. Bosch, N. S. de Carvalho, M. J. Germar, K. Peters, J. Paavonen, M. C. Bozonnat, D. Descamps, F. Struyf, G. O. Dubin, D. Rosillon, and L. Baril. Natural history of progression of HPV infection to cervical lesion or clearance: analysis of the control arm of the large, randomised PATRICIA study. *PLoS One*, 8(11):e79260, 2013.

[127] A. Giannoudis and C. S. Herrington. Human papillomavirus variants and squamous neoplasia of the cervix. *J Pathol*, 193(3):295–302, 2001.

[128] S. Marur, G. D'Souza, W. H. Westra, and A. A. Forastiere. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol*, 11(8):781–9, 2010.

[129] T. Ramqvist, N. Grun, and T. Dalianis. Human papillomavirus and tonsillar and base of tongue cancer. *Viruses*, 7(3):1332–43, 2015.

[130] L. Licitra, J. Bernier, C. Grandi, M. Merlano, P. Bruzzi, and J. L. Lefebvre. Cancer of the oropharynx. *Crit Rev Oncol Hematol*, 41(1):107–22, 2002.

[131] L. Marklund, A. Nasman, T. Ramqvist, T. Dalianis, E. Munck-Wikland, and L. Hammarstedt. Prevalence of human papillomavirus and survival in oropharyngeal cancer other than tonsil or base of tongue cancer. *Cancer Med*, 1(1):82–8, 2012.

[132] M. W. Lingen, W. Xiao, A. Schmitt, B. Jiang, R. Pickard, P. Kreinbrink, B. Perez-Ordonez, R. C. Jordan, and M. L. Gillison. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. *Oral Oncol*, 49(1):1–8, 2013.

[133] M. L. Gillison, G. D'Souza, W. Westra, E. Sugar, W. Xiao, S. Begum, and R. Viscidi. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst*, 100(6):407–20, 2008.

[134] S. M. Schwartz, J. R. Daling, D. R. Doody, G. C. Wipf, J. J. Carter, M. M. Madeleine, E. J. Mao, E. D. Fitzgibbons, S. Huang, A. M. Beckmann, J. K. McDougall, and D. A. Galloway. Oral cancer risk in relation to sexual history and evidence of human papillomavirus infection. *J Natl Cancer Inst*, 90(21):1626–36, 1998.

[135] A. K. Chaturvedi, G. D'Souza, M. L. Gillison, and H. A. Katki. Burden of HPV-positive oropharynx cancers among ever and never smokers in the U.S. population. *Oral Oncol*, 60:61–7, 2016.

[136] C. Fakhry, K. K. Andersen, J. Christensen, N. Agrawal, and D. W. Eisele. The Impact of Tonsillectomy upon the Risk of Oropharyngeal Carcinoma Diagnosis and Prognosis in the Danish Cancer Registry. *Cancer Prev Res (Phila)*, 8(7):583–9, 2015.

[137] A. K. Chaturvedi, H. Song, P. S. Rosenberg, T. Ramqvist, W. F. Anderson, E. Munck-Wikland, W. Ye, and T. Dalianis. Tonsillectomy and Incidence of Oropharyngeal Cancers. *Cancer Epidemiol Biomarkers Prev*, 25(6):944–50, 2016.

[138] C. Schnelle, D. C. Whiteman, S. V. Porceddu, B. J. Panizza, and A. Antonsson. Past sexual behaviors and risks of oropharyngeal squamous cell carcinoma: a case-case comparison. *Int J Cancer*, 140(5):1027–1034, 2017.

[139] J. D. Combes, A. A. Chen, and S. Franceschi. Prevalence of human papillomavirus in cancer of the oropharynx by gender. *Cancer Epidemiol Biomarkers Prev*, 23(12):2954–8, 2014.

[140] A. K. Chaturvedi, W. F. Anderson, J. Lortet-Tieulent, M. P. Curado, J. Ferlay, S. Franceschi, P. S. Rosenberg, F. Bray, and M. L. Gillison. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *J Clin Oncol*, 31(36):4550–9, 2013.

[141] M. Melbye and P. Sprogel. Aetiological parallel between anal cancer and cervical cancer. *Lancet*, 338(8768):657–9, 1991.

[142] M. Frisch, B. Glimelius, A. J. van den Brule, J. Wohlfahrt, C. J. Meijer, J. M. Walboomers, S. Goldman, C. Svensson, H. O. Adami, and M. Melbye. Sexually transmitted infection as a cause of anal cancer. *N Engl J Med*, 337(19):1350–8, 1997.

[143] B. E. Hoots, J. M. Palefsky, J. M. Pimenta, and J. S. Smith. Human papillomavirus type distribution in anal cancer and anal intraepithelial lesions. *Int J Cancer*, 124(10):2375–83, 2009.

[144] D. A. Machalek, M. Poynten, F. Jin, C. K. Fairley, A. Farnsworth, S. M. Garland, R. J. Hillman, K. Petoumenos, J. Roberts, S. N. Tabrizi, D. J. Templeton, and A. E. Grulich. Anal human papillomavirus infection and associated neoplastic lesions in men who have sex with men: a systematic review and meta-analysis. *Lancet Oncol*, 13(5):487–500, 2012.

[145] M. Frisch, E. Smith, A. Grulich, and C. Johansen. Cancer in a population-based cohort of men and women in registered homosexual partnerships. *Am J Epidemiol*, 157(11):966–72, 2003.

[146] J. R. Daling, M. M. Madeleine, L. G. Johnson, S. M. Schwartz, K. A. Shera, M. A. Wurscher, J. J. Carter, P. L. Porter, D. A. Galloway, and J. K. McDougall. Human papillomavirus, smoking, and sexual practices in the etiology of anal cancer. *Cancer*, 101(2):270–80, 2004.

[147] R. P. van der Zee, O. Richel, H. J. de Vries, and J. M. Prins. The increasing incidence of anal cancer: can it be explained by trends in risk groups? *Neth J Med*, 71(8):401–11, 2013.

[148] Epidemiology National Cancer Institute, Surveillance and End Results Program. Seer*explorer, 2019.

[149] N. Li, S. Franceschi, R. Howell-Jones, P. J. Snijders, and G. M. Clifford. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int J Cancer*, 128(4):927–35, 2011.

[150] S. M. Garland, W. Dimech, P. Collignon, L. Cooley, G. R. Nimmo, D. W. Smith, R. Baird, W. Rawlinson, A. M. Costa, and G. Higgins. The new screening program to prevent cervical cancer using HPV DNA: getting the balance right in maintaining quality. *J Pathol Clin Res*, 4(4):207–212, 2018.

[151] S. Arrossi, L. Thouyaret, R. Laudi, O. Marin, J. Ramirez, M. Paolino, R. Herrero, and A. Campanera. Implementation of HPV-testing for cervical cancer screening in programmatic contexts: The Jujuy demonstration project in Argentina. *Int J Cancer*, 137(7):1709–18, 2015.

[152] K. M. Elfstrom, L. Arnheim-Dahlstrom, L. von Karsa, and J. Dillner. Cervical cancer screening in Europe: Quality assurance and organisation of programmes. *Eur J Cancer*, 51(8):950–68, 2015.

[153] T. Haldorsen, G. B. Skare, G. Ursin, and T. Bjorge. Results of delayed triage by HPV testing and cytology in the Norwegian Cervical Cancer Screening Programme. *Acta Oncol*, 54(2):200–9, 2015.

[154] G. Koliopoulos, V. N. Nyaga, N. Santesso, A. Bryant, P. P. Martin-Hirsch, R. A. Mustafa, H. Schunemann, E. Paraskevaidis, and M. Arbyn. Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database Syst Rev*, 8:CD008587, 2017.

[155] N. Santesso, R. A. Mustafa, W. Wiercioch, R. Kehar, S. Gandhi, Y. Chen, A. Cheung, J. Hopkins, R. Khatib, B. Ma, A. A. Mustafa, N. Lloyd, D. Wu, N. Broutet, and H. J. Schunemann. Systematic reviews and meta-analyses of benefits and harms of cryotherapy, LEEP, and cold knife conization to treat cervical intraepithelial neoplasia. *Int J Gynaecol Obstet*, 132(3):266–71, 2016.

[156] O. Adegoke, S. Kulasingam, and B. Virnig. Cervical cancer trends in the United States: a 35-year population-based analysis. *J Womens Health (Larchmt)*, 21(10):1031–7, 2012.

[157] C. Fakhry, B. T. Rosenthal, D. P. Clark, and M. L. Gillison. Associations between oral HPV16 infection and cytopathology: evaluation of an oropharyngeal "pap-test equivalent" in high-risk populations. *Cancer Prev Res (Phila)*, 4(9):1378–84, 2011.

[158] P. A. Fox. Treatment options for anal intraepithelial neoplasia and evidence for their effectiveness. *Sex Health*, 9(6):587–92, 2012.

[159] J. Paavonen, P. Naud, J. Salmeron, C. M. Wheeler, S. N. Chow, D. Apter, H. Kitchener, X. Castellsague, J. C. Teixeira, S. R. Skinner, J. Hedrick, U. Jaisamrarn, G. Limson, S. Garland, A. Szarewski, B. Romanowski, F. Y. Aoki, T. F. Schwarz, W. A. Poppe, F. X. Bosch, D. Jenkins, K. Hardt, T. Zahaf, D. Descamps, F. Struyf, M. Lehtinen, and G. Dubin. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet*, 374(9686):301–14, 2009.

[160] S. M. Garland, M. Hernandez-Avila, C. M. Wheeler, G. Perez, D. M. Harper, S. Leodolter, G. W. Tang, D. G. Ferris, M. Steben, J. Bryan, F. J. Taddeo, R. Railkar, M. T. Esser, H. L. Sings, M. Nelson, J. Boslego, C. Sattler, E. Barr, and L. A. Koutsky. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. *N Engl J Med*, 356(19):1928–43, 2007.

[161] E. A. Joura, A. R. Giuliano, O. E. Iversen, C. Bouchard, C. Mao, J. Mehlsen, Jr. Moreira, E. D., Y. Ngan, L. K. Petersen, E. Lazcano-Ponce, P. Pitisuttithum, J. A. Restrepo, G. Stuart, L. Woelber, Y. C. Yang, J. Cuzick, S. M. Garland, W. Huh, S. K. Kjaer, O. M. Bautista, I. S. Chan, J. Chen, R. Gesser, E. Moeller, M. Ritter, S. Vuocolo, and A. Luxembourg. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med*, 372(8):711–23, 2015.

[162] Y. Delere, O. Wichmann, S. J. Klug, M. van der Sande, M. Terhardt, F. Zepp, and T. Harder. The efficacy and duration of vaccine protection against human papillomavirus: a systematic review and meta-analysis. *Dtsch Arztebl Int*, 111(35-36):584–91, 2014.

[163] T. F. Schwarz, A. Galaj, M. Spaczynski, J. Wysocki, A. M. Kaufmann, S. Poncelet, P. V. Suryakiran, N. Folschweiller, F. Thomas, L. Lin, and F. Struyf. Ten-year immune persistence and safety of the HPV-16/18 AS04-adjuvanted vaccine in females vaccinated at 15-55 years of age. *Cancer Med*, 6(11):2723–2731, 2017.

[164] D. C. Beachler, A. R. Kreimer, M. Schiffman, R. Herrero, S. Wacholder, A. C. Rodriguez, D. R. Lowy, C. Porras, J. T. Schiller, W. Quint, S. Jimenez, M. Safaeian, L. Struijk, J. Schussler, A. Hildesheim, and P. Gonzalez. Multisite HPV16/18 Vaccine Efficacy Against Cervical, Anal, and Oral HPV Infection. *J Natl Cancer Inst*, 108(1), 2016.

[165] A. R. Giuliano, J. M. Palefsky, S. Goldstone, Jr. Moreira, E. D., M. E. Penny, C. Aranda, E. Vardas, H. Moi, H. Jessen, R. Hillman, Y. H. Chang, D. Ferris, D. Rouleau, J. Bryan, J. B. Marshall, S. Vuocolo, E. Barr, D. Radley, R. M. Haupt, and D. Guris. Efficacy of quadrivalent HPV vaccine against HPV Infection and disease in males. *N Engl J Med*, 364(5):401–11, 2011.

[166] L. E. Markowitz, V. Tsu, S. L. Deeks, H. Cubie, S. A. Wang, A. S. Vicari, and J. M. Brotherton. Human papillomavirus vaccine introduction–the first five years. *Vaccine*, 30 Suppl 5:F139–48, 2012.

[167] R. Aggarwal, AJ. Pollard, N. Bhatla, S. Franceschi, and E. L. Franco. Sage working group on human papillomavirus immunization. 2018.

[168] M. E. Halloran and C. J. Struchiner. Causal inference in infectious diseases. *Epidemiology*, 6(2):142–51, 1995.

[169] M. Brisson, E. Benard, M. Drolet, J. A. Bogaards, I. Baussano, S. Vanska, M. Jit, M. C. Boily, M. A. Smith, J. Berkhof, K. Canfell, H. W. Chesson, E. A. Burger, Y. H. Choi, B. F. De Blasio, S. J. De Vlas, G. Guzzetta, J. A. C. Hontelez, J. Horn, M. R. Jepsen, J. J. Kim, F. Lazzarato, S. M. Matthijsse, R. Mikolajczyk, A. Pavelyev, M. Pillsbury, L. A. Shafer, S. P. Tully, H. C. Turner, C. Usher, and C. Walsh. Population-level impact, herd immunity, and elimination after human papillomavirus vaccination: a systematic review and meta-analysis of predictions from transmission-dynamic models. *Lancet Public Health*, 1(1), 2016.

[170] M. Drolet, E. Benard, M. Jit, R. Hutubessy, and M. Brisson. Model comparisons of the effectiveness and cost-effectiveness of vaccination: A systematic review of the literature. *Value Health*, 21(10):1250–1258, 2018.

[171] S. S. Ng, R. Hutubessy, and N. Chaiyakunapruk. Systematic review of cost-effectiveness studies of human papillomavirus (HPV) vaccination: 9-Valent vaccine, gender-neutral and multiple age cohort vaccination. *Vaccine*, 36(19):2529–2544, 2018.

[172] M. Drolet, J. F. Laprise, M. C. Boily, E. L. Franco, and M. Brisson. Potential cost-effectiveness of the nonavalent human papillomavirus (HPV) vaccine. *Int J Cancer*, 134(9):2264–8, 2014.

[173] J. F. Laprise, L. E. Markowitz, H. W. Chesson, M. Drolet, and M. Brisson. Comparison of 2-Dose and 3-Dose 9-Valent Human Papillomavirus Vaccine Schedules in the United States: A Cost-effectiveness Analysis. *J Infect Dis*, 214(5):685–8, 2016.

[174] M. Jit, M. Brisson, J. F. Laprise, and Y. H. Choi. Comparison of two dose and three dose human papillomavirus vaccine schedules: cost effectiveness analysis based on transmission model. *BMJ*, 350:g7584, 2015.

[175] M. Jit, J. F. Laprise, Y. H. Choi, and M. Brisson. Fewer than three doses of HPV vaccine. *Lancet Oncol*, 16(9):e423–e424, 2015.

[176] I. Baussano, G. Ronco, N. Segnan, K. French, P. Vineis, and G. P. Garnett. HPV-16 infection and cervical cancer: modeling the influence of duration of infection and precancerous lesions. *Epidemics*, 2(1):21–8, 2010.

[177] E. H. Elbasha, E. J. Dasbach, and R. P. Insinga. Model for assessing human papillomavirus vaccination strategies. *Emerg Infect Dis*, 13(1):28–41, 2007.

[178] J. A. Bogaards, M. Xiridou, V. M. Coupe, C. J. Meijer, J. Wallinga, and J. Berkhof. 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*, 171(7):817–25, 2010.

[179] I. A. Korostil, H. Ali, R. J. Guy, B. Donovan, M. G. Law, and D. G. Regan. Near elimination of genital warts in Australia predicted with extension of human papillomavirus vaccination to males. *Sex Transm Dis*, 40(11):833–5, 2013.

[180] N. Van de Velde, M. Brisson, and M. C. Boily. Understanding differences in predictions of HPV vaccine effectiveness: A comparative model-based analysis. *Vaccine*, 28(33):5473–84, 2010.

[181] H. W. Chesson, E. Meites, D. U. Ekwueme, M. Saraiya, and L. E. Markowitz. Cost-effectiveness of nonavalent HPV vaccination among males aged 22 through 26years in the United States. *Vaccine*, 36(29):4362–4368, 2018.

[182] M. Brisson, N. Van de Velde, P. De Wals, and M. C. Boily. The potential cost-effectiveness of prophylactic human papillomavirus vaccines in Canada. *Vaccine*, 25(29):5399–408, 2007.

[183] W. J. Edmunds, C. J. O'Callaghan, and D. J. Nokes. Who mixes with whom? a method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc Biol Sci*, 264(1384):949–57, 1997.

[184] J. J. Kim, M. Brisson, W. J. Edmunds, and S. J. Goldie. Modeling cervical cancer prevention in developed countries. *Vaccine*, 26 Suppl 10:K76–86, 2008.

[185] J. D. Goldhaber-Fiebert, N. K. Stout, J. Ortendahl, K. M. Kuntz, S. J. Goldie, and J. A. Salomon. Modeling human papillomavirus and cervical cancer in the United States for analyses of screening and vaccination. *Popul Health Metr*, 5:11, 2007.

[186] N. Van de Velde, M. C. Boily, M. Drolet, E. L. Franco, M. H. Mayrand, E. V. Kliewer, F. Coutlee, J. F. Laprise, T. Malagon, and M. Brisson. Population-level impact of the bivalent, quadrivalent, and nonavalent human papillomavirus vaccines: a model-based analysis. *J Natl Cancer Inst*, 104(22):1712–23, 2012.

[187] M. Jit, Y. H. Choi, and W. J. Edmunds. Economic evaluation of human papillomavirus vaccination in the United Kingdom. *BMJ*, 337:a769, 2008.

[188] K. Canfell, H. Chesson, S. L. Kulasingam, J. Berkhof, M. Diaz, and J. J. Kim. Modeling preventative strategies against human papillomavirus-related disease in developed countries. *Vaccine*, 30 Suppl 5:F157–67, 2012.

[189] R. A. Borracci, S. V. Segal, and J. H. Mendez. Epidemiological dynamic modeling of human papillomavirus-related diseases to assess vaccination strategies in Argentina. *Medicina (B Aires)*, 78(5):315–328, 2018.

[190] A. F. Brouwer, R. Meza, and M. C. Eisenberg. Transmission heterogeneity and autoinoculation in a multisite infection model of HPV. *Math Biosci*, 270(Pt A):115–25, 2015.

[191] S. Alizon, C. L. Murall, and I. G. Bravo. Why Human Papillomavirus Acute Infections Matter. *Viruses*, 9(10), 2017.

[192] M. Hofler. Causal inference based on counterfactuals. *BMC Med Res Methodol*, 5:28, 2005.

[193] A. Camacho, R. M. Eggo, S. Funk, C. H. Watson, A. J. Kucharski, and W. J. Edmunds. Estimating the probability of demonstrating vaccine efficacy in the declining ebola epidemic: a bayesian modelling approach. *BMJ Open*, 5(12):e009346, 2015.

[194] M. D. Hitchings, R. F. Grais, and M. Lipsitch. Using simulation to aid trial design: Ring-vaccination trials. *PLoS Negl Trop Dis*, 11(3):e0005470, 2017.

[195] M. C. Boily, B. Masse, R. Alsallaq, N. S. Padian, J. W. Eaton, J. F. Vesga, and T. B. Hallett. HIV treatment as prevention: considerations in the design, conduct, and analysis of cluster randomized controlled trials of combination HIV prevention. *PLoS Med*, 9(7):e1001250, 2012.

[196] M. L. Gillison, L. Alemany, P. J. Snijders, A. Chaturvedi, B. M. Steinberg, S. Schwartz, and X. Castellsague. Human papillomavirus and diseases of the upper airway: head and neck cancer and respiratory papillomatosis. *Vaccine*, 30 Suppl 5:F34–54, 2012.

[197] N. Munoz, S. K. Kjaer, K. Sigurdsson, O. E. Iversen, M. Hernandez-Avila, C. M. Wheeler, G. Perez, D. R. Brown, L. A. Koutsky, E. H. Tay, P. J. Garcia, K. A. Ault, S. M. Garland, S. Leodolter, S. E. Olsson, G. W. Tang, D. G. Ferris, J. Paavonen, M. Steben, F. X. Bosch, J. Dillner, W. K. Huh, E. A. Joura, R. J. Kurman, S. Majewski, E. R. Myers, L. L. Villa, F. J. Taddeo, C. Roberts, A. Tadesse, J. T. Bryan, L. C. Lupinacci, K. E. Giacoletti, H. L. Sings, M. K. James, T. M. Hesley, E. Barr, and R. M. Haupt. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. *J Natl Cancer Inst*, 102(5):325–39, 2010.

[198] S. E. Goldstone, H. Jessen, J. M. Palefsky, A. R. Giuliano, Jr. Moreira, E. D., E. Vardas, C. Aranda, R. J. Hillman, D. G. Ferris, F. Coutlee, J. B. Marshall, S. Vuocolo, R. M. Haupt, D. Guris, and E. Garner. Quadrivalent HPV vaccine efficacy against disease related to vaccine and non-vaccine HPV types in males. *Vaccine*, 31(37):3849–55, 2013.

[199] R. Herrero, W. Quint, A. Hildesheim, P. Gonzalez, L. Struijk, H. A. Katki, C. Porras, M. Schiffman, A. C. Rodriguez, D. Solomon, S. Jimenez, J. T. Schiller, D. R. Lowy, L. J. van Doorn, S. Wacholder, and A. R. Kreimer. Reduced prevalence of oral human papillomavirus (HPV) 4 years after bivalent HPV vaccination in a randomized clinical trial in Costa Rica. *PLoS One*, 8(7):e68329, 2013.

[200] M. L. Gillison, X. Castellsague, A. Chaturvedi, M. T. Goodman, P. Snijders, M. Tommasino, M. Arbyn, and S. Franceschi. Eurogin Roadmap: comparative epidemiology of HPV infection and associated cancers of the head and neck and cervix. *Int J Cancer*, 134(3):497–507, 2014.

[201] S. L. Vogt, P. E. Gravitt, N. A. Martinson, J. Hoffmann, and G. D'Souza. Concordant oral-genital HPV infection in South Africa couples: Evidence for transmission. *Front Oncol*, 3:303, 2013.

[202] M. Brisson, É. Bénard, M. Drolet, I. Baussano, and J. Berkhof. Population-level impact, herd immunity and elimination after HPV vaccination: a systemic review and meta-analysis of predictions of 17 transmission-dynamic models, 2015.

[203] M. Jit and M. Brisson. Modelling the epidemiology of infectious diseases for decision analysis: a primer. *Pharmacoeconomics*, 29(5):371–86, 2011.

[204] S. Hariri, E. R. Unger, M. Sternberg, E. F. Dunne, D. Swan, S. Patel, and L. E. Markowitz. Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006. *J Infect Dis*, 204(4):566–73, 2011.

[205] M. T. Goodman, Y. B. Shvetsov, K. McDuffie, L. R. Wilkens, X. Zhu, L. Ning, J. Killeen, L. Kamemoto, and B. Y. Hernandez. Acquisition of anal human papillomavirus (HPV) infection in women: the Hawaii HPV Cohort study. *J Infect Dis*, 197(7):957–66, 2008.

[206] R. P. Insinga, E. J. Dasbach, E. H. Elbasha, K. L. Liaw, and E. Barr. Incidence and duration of cervical human papillomavirus 6, 11, 16, and 18 infections in young women: an evaluation from multiple analytic perspectives. *Cancer Epidemiol Biomarkers Prev*, 16(4):709–15, 2007.

[207] M. Brisson, J. F. Laprise, M. Drolet, N. Van de Velde, E. L. Franco, E. V. Kliewer, G. Ogilvie, S. L. Deeks, and M. C. Boily. Comparative cost-effectiveness of the quadrivalent and bivalent human papillomavirus vaccines: a transmission-dynamic modeling study. *Vaccine*, 31(37):3863–71, 2013.

[208] M. D. McKay, R. J. Beckman, and W. J. Conover. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, 1979.

[209] P. Driessche and James Watmough. *Further Notes on the Basic Reproduction Number*, pages 159–178. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[210] A. F. Brouwer, M. C. Eisenberg, T. E. Carey, and R. Meza. Trends in HPV cervical and seroprevalence and associations between oral and genital infection and serum antibodies in NHANES 2003-2012. *BMC Infect Dis*, 15:575, 2015.

[211] T. Gebhardt and L. K. Mackay. Local immunity by tissue-resident CD8(+) memory T cells. *Front Immunol*, 3:340, 2012.

[212] D. C. Beachler, R. Viscidi, E. A. Sugar, H. Minkoff, H. D. Strickler, R. D. Cranston, D. J. Wiley, L. P. Jacobson, K. M. Weber, J. B. Margolick, S. Reddy, M. L. Gillison, and G. D'Souza. A longitudinal study of human papillomavirus 16 L1, e6, and e7 seropositivity and oral human papillomavirus 16 infection. *Sex Transm Dis*, 42(2):93–7, 2015.

[213] C. M. Pierce Campbell, R. P. Viscidi, B. N. Torres, H. Y. Lin, W. Fulp, M. Abrahamsen, E. Lazcano-Ponce, L. L. Villa, A. R. Kreimer, and A. R. Giuliano. Human Papillomavirus (HPV) L1 Serum Antibodies and the Risk of Subsequent Oral HPV Acquisition in Men: The HIM Study. *J Infect Dis*, 2016.

[214] J. C. M. Heijne, Gafs van Liere, Cjpa Hoebe, J. A. Bogaards, B. H. B. van Benthem, and Nhtm Dukers-Muijrers. What explains anorectal chlamydia infection in women? implications of a mathematical model for test and treatment strategies. *Sex Transm Infect*, 93(4):270–275, 2017.

[215] B. Hui, C. K. Fairley, M. Chen, A. Grulich, J. Hocking, G. Prestage, S. Walker, M. Law, and D. Regan. Oral and anal sex are key to sustaining gonorrhoea at endemic levels in MSM populations: a mathematical model. *Sex Transm Infect*, 91(5):365–9, 2015.

[216] R. L. Moy, Y. D. Eliezri, G. J. Nuovo, J. A. Zitelli, R. G. Bennett, and S. Silverstein. Human papillomavirus type 16 DNA in periungual squamous cell carcinomas. *JAMA*, 261(18):2669–73, 1989.

[217] M. D. Ryser, A. Rositch, and P. E. Gravitt. Modeling of US Human Papillomavirus (HPV) Seroprevalence by Age and Sexual Behavior Indicates an Increasing Trend of HPV Infection Following the Sexual Revolution. *J Infect Dis*, 216(5):604–611, 2017.

[218] S. Vaccarella, J. Lortet-Tieulent, M. Plummer, S. Franceschi, and F. Bray. Worldwide trends in cervical cancer incidence: impact of screening against changes in disease risk factors. *Eur J Cancer*, 49(15):3262–73, 2013.

[219] L. Xu, K. R. Dahlstrom, D. R. Lairson, and E. M. Sturgis. Projected oropharyngeal carcinoma incidence among middle-aged US men. *Head Neck*, 2019.

[220] M. Jit, R. Chapman, O. Hughes, and Y. H. Choi. Comparing bivalent and quadrivalent human papillomavirus vaccines: economic evaluation based on transmission model. *BMJ*, 343:d5775, 2011.

[221] H. W. Chesson, J. F. Laprise, M. Brisson, and L. E. Markowitz. Impact and Cost-effectiveness of 3 Doses of 9-Valent Human Papillomavirus (HPV) Vaccine Among US Females Previously Vaccinated With 4-Valent HPV Vaccine. *J Infect Dis*, 213(11):1694–700, 2016.

[222] M. Brisson, J. F. Laprise, H. W. Chesson, M. Drolet, T. Malagon, M. C. Boily, and L. E. Markowitz. Health and Economic Impact of Switching from a 4-Valent to a 9-Valent HPV Vaccination Program in the United States. *J Natl Cancer Inst*, 108(1), 2016.

[223] G. M. Clifford, J. S. Smith, T. Aguado, and S. Franceschi. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. *Br J Cancer*, 89(1):101–5, 2003.

[224] CDC. Table 1. resident population, by age, sex, race, and hispanic origin: United states, selected years 1950-2012, 2012.

[225] L. T. Haderxhanaj, J. S. Leichliter, S. O. Aral, and H. W. Chesson. Sex in a lifetime: Sexual behaviors in the united states by lifetime number of sex partners, 2006-2010. *Sex Transm Dis*, 41(6):345–52, 2014.

[226] M. Zappa, C. B. Visioli, S. Ciatto, A. Iossa, E. Paci, and P. Sasieni. Lower protection of cytological screening for adenocarcinomas and shorter protection for younger women: the results of a case-control study in Florence. *Br J Cancer*, 90(9):1784–6, 2004.

[227] H. Mitchell, G. Medley, I. Gordon, and G. Giles. Cervical cytology reported as negative and risk of adenocarcinoma of the cervix: no strong evidence of benefit. *Br J Cancer*, 71(4):894–7, 1995.

[228] P. Laukkanen, P. Koskela, E. Pukkala, J. Dillner, E. Laara, P. Knekt, and M. Lehtinen. Time trends in incidence and prevalence of human papillomavirus type 6, 11 and 16 infections in finland. *J Gen Virol*, 84(Pt 8):2105–9, 2003.

[229] V. af Geijersstam, Z. Wang, I. Lewensohn-Fuchs, C. Eklund, J. T. Schiller, M. Forsgren, and J. Dillner. Trends in seroprevalence of human papillomavirus type 16 among pregnant women in stockholm, sweden, during 1969-1989. *Int J Cancer*, 76(3):341–4, 1998.

[230] A. A. Adimora and V. J. Schoenbach. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *J Infect Dis*, 191 Suppl 1:S115–22, 2005.

[231] P. Fine, K. Eames, and D. L. Heymann. "herd immunity": a rough guide. *Clin Infect Dis*, 52(7):911–6, 2011.

[232] N. F. Schlecht, E. L. Franco, T. E. Rohan, S. K. Kjaer, M. H. Schiffman, A. B. Moscicki, and S. W. Duffy. Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews. *J Epidemiol Biostat*, 6(5):393–407, 2001.

[233] A. S. Furber, R. Maheswaran, J. N. Newell, and C. Carroll. Is smoking tobacco an independent risk factor for HIV infection and progression to AIDS? A systemic review. *Sex Transm Infect*, 83(1):41–6, 2007.

[234] I. A. Doherty, N. S. Padian, C. Marlow, and S. O. Aral. Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *J Infect Dis*, 191 Suppl 1:S42–54, 2005.

[235] A. Agrawal, A. C. Heath, J. D. Grant, M. L. Pergadia, D. J. Statham, K. K. Bucholz, N. G. Martin, and P. A. Madden. Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behav Genet*, 36(4):553–66, 2006.

[236] A. E. Clark and F. Etile. Don't give up on me baby: spousal correlation in smoking behaviour. *J Health Econ*, 25(5):958–78, 2006.

[237] S. Vaccarella, R. Herrero, P. J. Snijders, M. Dai, J. O. Thomas, N. T. Hieu, C. Ferreccio, E. Matos, H. Posso, S. de Sanjose, H. R. Shin, S. Sukvirach, E. Lazcano-Ponce, N. Munoz, C. J. Meijer, and S. Franceschi. Smoking and human papillomavirus infection: pooled analysis of the International Agency for Research on cancer HPV prevalence surveys. *Int J Epidemiol*, 37(3):536–46, 2008.

[238] P. A. Cavazos-Rehg, M. J. Krauss, E. L. Spitznagel, M. Schootman, L. B. Cottler, and L. J. Bierut. Number of sexual partners and associations with initiation and intensity of substance use. *AIDS Behav*, 15(4):869–74, 2011.

[239] M. C. Boily and R. M. Anderson. Human immunodeficiency virus transmission and the role of other sexually transmitted diseases. Measures of association and study design. *Sex Transm Dis*, 23(4):312–32, 1996.

[240] K. N. Desai, M. C. Boily, B. R. Masse, M. Alary, and R. M. Anderson. Simulation studies of phase III clinical trials to test the efficacy of a candidate HIV-1 vaccine. *Epidemiol Infect*, 123(1):65–88, 1999.

[241] M. T. White, J. T. Griffin, C. J. Drakeley, and A. C. Ghani. Heterogeneity in malaria exposure and vaccine response: implications for the interpretation of vaccine efficacy trials. *Malar J*, 9:82, 2010.

[242] J. S. Koopman and J. W. Lynch. Individual causal models and population system models in epidemiology. *Am J Public Health*, 89(8):1170–4, 1999.

[243] R. Wolf and D. Freedman. Cigarette smoking, sexually transmitted diseases, and HIV/AIDS. *Int J Dermatol*, 39(1):1–9, 2000.

[244] D. C. Beachler, K. M. Weber, J. B. Margolick, H. D. Strickler, R. D. Cranston, R. D. Burk, D. J. Wiley, H. Minkoff, S. Reddy, E. E. Stammer, M. L. Gillison, and G. D'Souza. Risk factors for oral HPV infection among a high prevalence population of HIV-positive and at-risk HIV-negative adults. *Cancer Epidemiology, Biomarkers and Prevention*, 21(1):122–33, 2012.

[245] J. W. Sellors, J. B. Mahony, J. Kaczorowski, A. Lytwyn, H. Bangura, S. Chong, A. Lorincz, D. M. Dalby, V. Janjusevic, and J. L. Keller. Prevalence and predictors of human papillomavirus infection in women in Ontario, Canada. Survey of HPV in Ontario Women (SHOW) Group. *CMAJ*, 163(5):503–8, 2000.

[246] F. Granath, J. Giesecke, G. Scalia-Tomba, K. Ramstedt, and L. Forssman. 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*, 107(2):341–8, 1991.

[247] S. O. Aral, J. P. Hughes, B. Stoner, W. Whittington, H. H. Handsfield, R. M. Anderson, and K. K. Holmes. Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *Am J Public Health*, 89(6):825–33, 1999.

[248] H. C. Johnson, K. M. Elfstrom, and W. J. Edmunds. Inference of type-specific HPV transmissibility, progression and clearance rates: a mathematical modelling approach. *PLoS One*, 7(11):e49614, 2012.

[249] M. Drolet, M. Brisson, E. Maunsell, E. L. Franco, F. Coutlee, A. Ferenczy, W. Fisher, and J. A. Mansi. The psychosocial impact of an abnormal cervical smear result. *Psychooncology*, 21(10):1071–81, 2012.

[250] L. E. Manhart, S. O. Aral, K. K. Holmes, C. W. Critchlow, J. P. Hughes, W. L. Whittington, and B. Foxman. 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*, 160(4):393–402, 2004.

[251] F. Liljeros, C. R. Edling, and L. A. Nunes Amaral. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes Infect*, 5(2):189–96, 2003.

[252] S. K. Kjaer, A. J. van den Brule, J. E. Bock, P. A. Poll, G. Engholm, M. E. Sherman, J. M. Walboomers, and C. J. Meijer. 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*, 6(10):799–805, 1997.

[253] E. L. Franco and A. R. Spence. Commentary: Smoking and human papillomavirus infection: the pursuit of credibility for an epidemiologic association. *Int J Epidemiol*, 37(3):547–8, 2008.

[254] A. M. Johnson, C. H. Mercer, B. Erens, A. J. Copas, S. McManus, K. Wellings, K. A. Fenton, C. Korovessis, W. Macdowall, K. Nanchahal, S. Purdon, and J. Field. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet*, 358(9296):1835–42, 2001.

[255] R. Hiscock, L. Bauld, A. Amos, J. A. Fidler, and M. Munafo. Socioeconomic status and smoking: a review. *Ann N Y Acad Sci*, 1248:107–23, 2012.

[256] E.O. Laumann, J.H. Gagnon, Michael R.T., and S. Michaels. *The social organization of sexuality.* The University of Chicago Press, Chicago, 1994.

[257] M. S. Cohen, Y. Q. Chen, M. McCauley, T. Gamble, M. C. Hosseinipour, N. Kumarasamy, J. G. Hakim, J. Kumwenda, B. Grinsztejn, J. H. Pilotto, S. V. Godbole, S. Mehendale, S. Chariyalertsak, B. R. Santos, K. H. Mayer, I. F. Hoffman, S. H. Eshleman, E. Piwowar-Manning, L. Wang, J. Makhema, L. A. Mills, G. de Bruyn, I. Sanne, J. Eron, J. Gallant, D. Havlir, S. Swindells, H. Ribaudo, V. Elharrar, D. Burns, T. E. Taha, K. Nielsen-Saines, D. Celentano, M. Essex, and T. R. Fleming. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*, 365(6):493–505, 2011.

[258] Y. J. Zhang, X. X. Feng, Y. G. Fan, Z. Y. Jiang, X. H. Zhong, M. Q. Li, and D. Q. Ye. HIV transmission and related risk factors among serodiscordant couples in Liuzhou, China. *J Med Virol*, 87(4):553–6, 2015.

[259] K. Kero, J. Rautava, K. Louvanto, K. Syrjanen, S. Grenman, and S. Syrjanen. Genotype-specific concordance of oral and genital human papillomavirus infections among marital couples is low. *Eur J Clin Microbiol Infect Dis*, 35(4):697–704, 2016.

[260] G. D'Souza, K. Cullen, J. Bowie, R. Thorpe, and C. Fakhry. Differences in oral sexual behaviors by gender, age, and race explain observed differences in prevalence of oral human papillomavirus infection. *PLoS One*, 9(1):e86023, 2014.

[261] D. X. Yang, P. R. Soulos, B. Davis, C. P. Gross, and J. B. Yu. Impact of widespread cervical cancer screening: Number of cancers prevented and changes in race-specific incidence. *Am J Clin Oncol*, 41(3):289–294, 2018.

[262] S. Vaccarella, S. Franceschi, G. Engholm, S. Lonnberg, S. Khan, and F. Bray. 50 years of screening in the Nordic countries: quantifying the effects on cervical cancer incidence. *Br J Cancer*, 111(5):965–9, 2014.

[263] L. E. Markowitz, M. Drolet, N. Perez, M. Jit, and M. Brisson. Human papillomavirus vaccine effectiveness by number of doses: Systematic review of data from national immunization programs. *Vaccine*, 36(32 Pt A):4806–4815, 2018.

[264] M. C. Fiore. Trends in cigarette smoking in the united states. the epidemiology of tobacco use. *Med Clin North Am*, 76(2):289–303, 1992.

[265] T. K. Greenfield and W. C. Kerr. Tracking alcohol consumption over time. *Alcohol Res Health*, 27(1):30–8, 2003.

[266] J. E. Darroch. Trends in contraceptive use. *Contraception*, 87(3):259–63, 2013.

[267] T. Malagon, A. N. Burchell, M. El-Zein, J. Guenoun, P. P. Tellier, F. Coutlee, and E. L. Franco. Estimating HPV DNA Deposition Between Sexual Partners Using HPV Concordance, Y Chromosome DNA Detection, and Self-reported Sexual Behaviors. *J Infect Dis*, 216(10):1210–1218, 2017.

[268] P. Lemieux-Mellouki, M. Drolet, J. Brisson, E. L. Franco, M. C. Boily, I. Baussano, and M. Brisson. Assortative mixing as a source of bias in epidemiological studies of sexually transmitted infections: the case of smoking and human papillomavirus. *Epidemiol Infect*, pages 1–10, 2015.

[269] U. John, M. Hanke, C. Meyer, and A. Schumann. Gender and age differences among current smokers in a general population survey. *BMC Public Health*, 5:57, 2005.

[270] H. Shigeishi and M. Sugiyama. Risk factors for oral human papillomavirus infection in healthy individuals: A systematic review and meta-analysis. *J Clin Med Res*, 8(10):721–9, 2016.

[271] M. I. Rodriguez-Alvarez, J. L. Gomez-Urquiza, H. Husein-El Ahmed, L. Albendin-Garcia, J. Gomez-Salgado, and G. A. Canadas-De la Fuente. Prevalence and risk factors of human papillomavirus in male patients: A systematic review and meta-analysis. *Int J Environ Res Public Health*, 15(10), 2018.

[272] National survey of family growth 2011-2013, 2015.

[273] H. W. Hethcote and H. R. Thieme. Stability of the endemic equilibrium in epidemic models with subpopulations. *Math. Biosci.*, 75:205–27, 1985.

[274] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, 28(4):365–382, 1990.

Appendix A

Supplementary materials. Modelling multi-site transmission of the human papillomavirus and its impact on vaccination effectiveness

A.1 (Objective 1): Comparing predictions of HPV16 vaccination effectiveness between multi-site and uni-site transmission-dynamic models

A.1.1 Models flow charts for the multi-site model

The Figures A.1 to A.4 present the models flow charts (deaths/births and gender/sexual activity stratifications are not represented) for the multi-site model used to compare predictions of HPV vaccination effectiveness between multi-site and uni-site transmission-dynamic models (Objective 1).

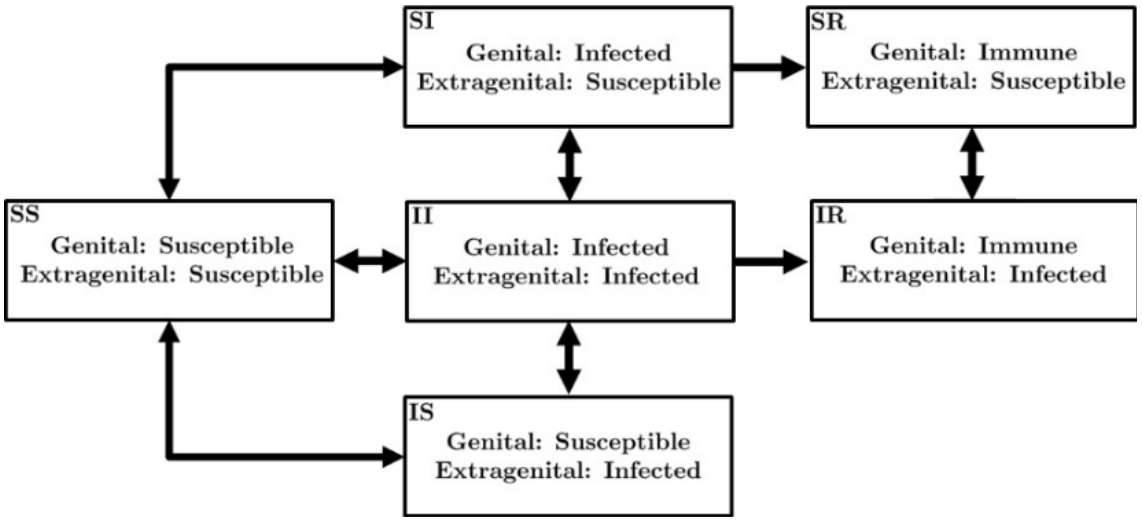


Figure A.1 – **Flow chart of multi-site model that include local immunity after genital infection (Scenario 1):** individuals can only acquire immunity upon clearing genital infection and immunity protects against subsequent genital infections.

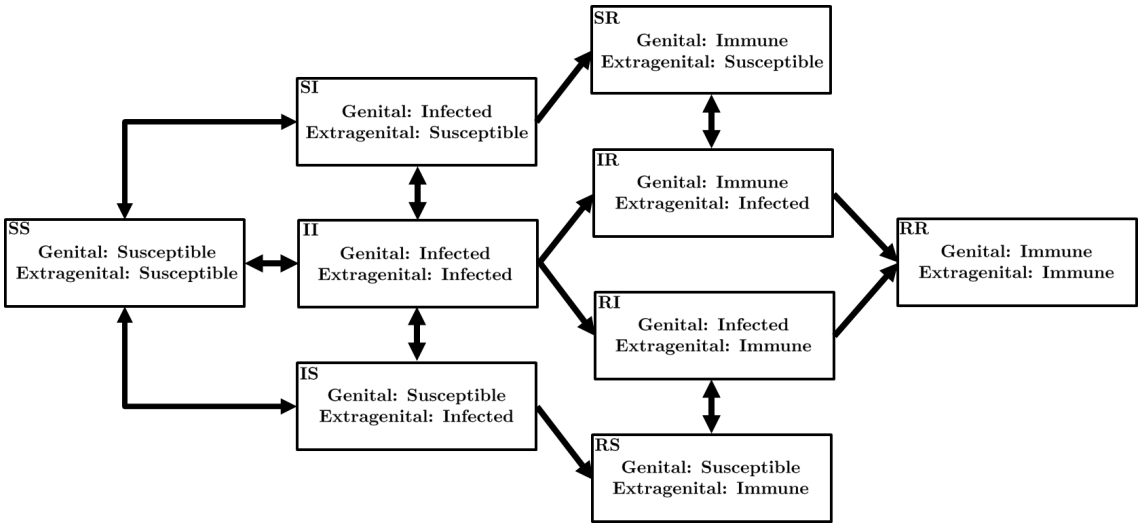


Figure A.2 – **Flow chart of multi-site model that include local immunity after genital and extra-genital infections (Scenario 2):** individuals can acquire local immunity upon clearing genital and extragenital infections.

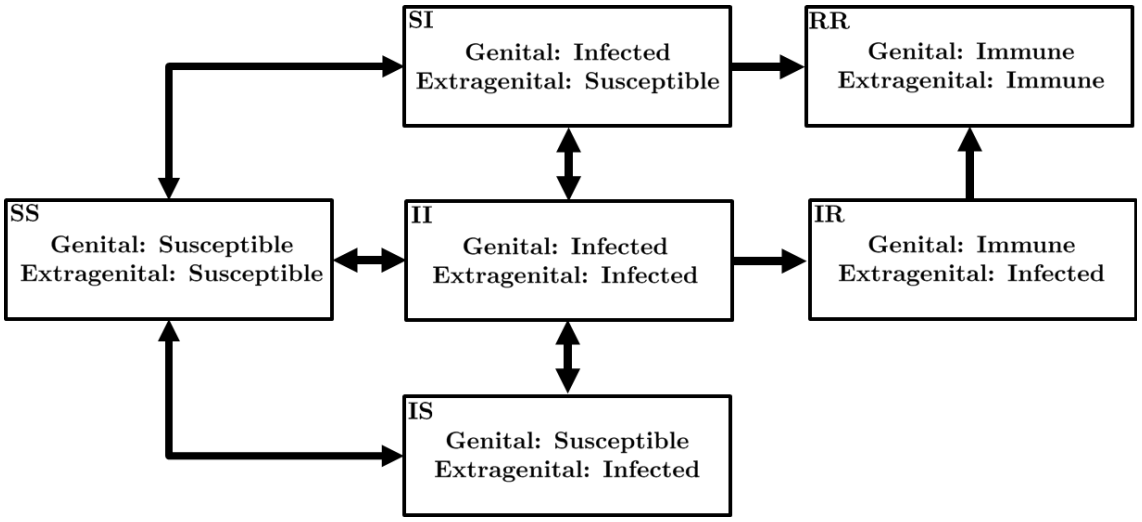


Figure A.3 – **Flow chart of multi-site model that include systemic immunity after genital infection (Scenario 3):** individuals can only acquire immunity upon clearing genital infection and immunity protects against subsequent infection at any site.

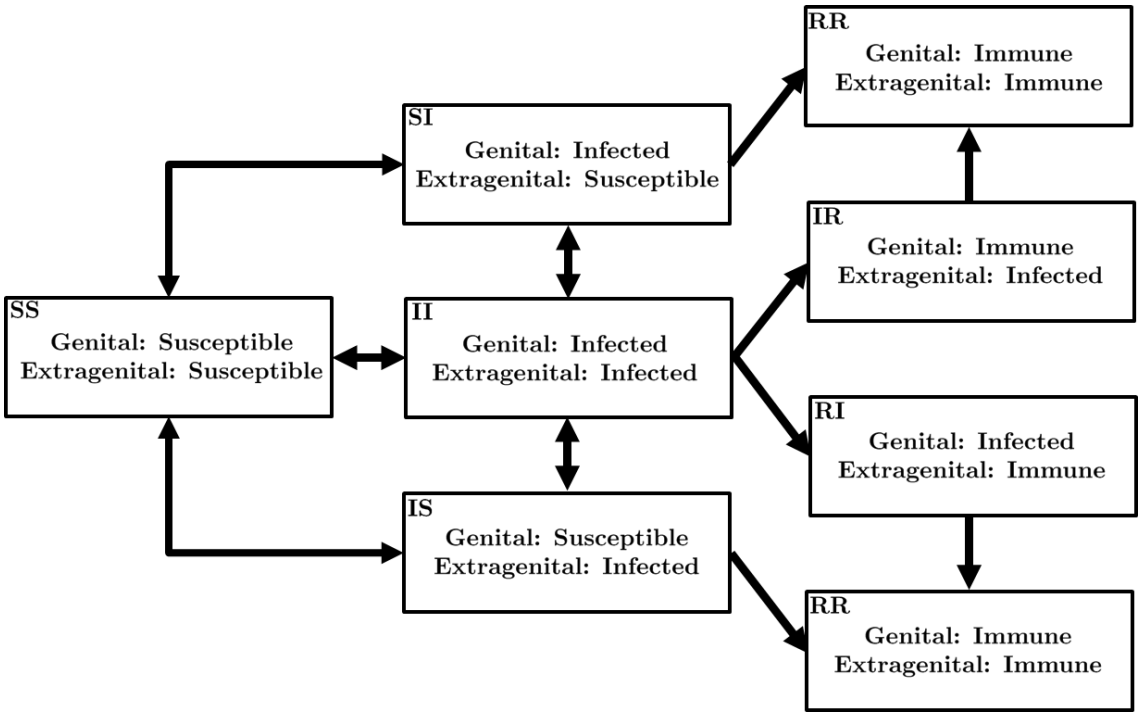


Figure A.4 – **Flow chart of multi-site model that include systemic immunity after genital and extragenital infections (Scenario 4):** individuals can acquire systemic immunity upon clearing genital or extragenital infection.

A.1.2 Model equations for local immunity

The set of differential equations A.1-A.9 is used to investigate the effect of multi-site transmission on vaccination impact assuming local immunity only after genital infection (scenario 1) or local immunity after genital infection and extragenital infection (scenario 2).

$$\begin{aligned}\frac{d}{dt}SI_{W,k} = \; & -\sigma_G SI_{W,k} - \mu SI_{W,k} + \sigma_E(1-\omega_E)II_{W,k} + c_k SS_{W,k}\left(\pi_E^G I_M^E + \pi_G^G I_M^G + \pi_{EG}^G I_M^{EG}\right) \\ & -c_k SI_{W,k}\left(\left(\pi_E^E + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^E + \pi_{EG}^{EG}\right) I_M^{EG}\right) - \lambda_G SI_{W,k}\end{aligned} \tag{A.1}$$

$$\begin{aligned}\frac{d}{dt}IS_{W,k} = \; & -\sigma_E IS_{W,k} - \mu IS_{W,k} + \sigma_G(1-\omega_G)II_{W,k} + c_k SS_{W,k}\left(\pi_E^E I_M^E + \pi_G^E I_M^G + \pi_{EG}^E I_M^{EG}\right) \\ & -c_k IS_{W,k}\left(\left(\pi_E^G + \pi_E^{EG}\right) I_M^E + \left(\pi_G^G + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^G + \pi_{EG}^{EG}\right) I_M^{EG}\right) - \lambda_E IS_{W,k}\end{aligned} \tag{A.2}$$

$$\begin{aligned}\frac{d}{dt}II_{W,k} = \; & -\sigma_E II_{W,k} - \mu II_{W,k} - \sigma_G II_{W,k} + c_k SS_{W,k}\left(\pi_E^{EG} I_M^E + \pi_G^{EG} I_M^G + \pi_{EG}^{EG} I_M^{EG}\right) \\ & +c_k SI_{W,k}\left(\left(\pi_E^E + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^E + \pi_{EG}^{EG}\right) I_M^{EG}\right) + \lambda_E IS_{W,k} \\ & +c_k IS_{W,k}\left(\left(\pi_E^G + \pi_E^{EG}\right) I_M^E + \left(\pi_G^G + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^G + \pi_{EG}^{EG}\right) I_M^{EG}\right) + \lambda_G SI_{W,k}\end{aligned} \tag{A.3}$$

$$\begin{aligned}\frac{d}{dt}IR_{W,k} = \; & -\sigma_E IR_{W,k} - \mu IR_{W,k} + \sigma_G \omega_G II_{W,k} \\ & +c_k SR_{W,k}\left(\left(\pi_E^E + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^E + \pi_{EG}^{EG}\right) I_M^{EG}\right)\end{aligned} \tag{A.4}$$

$$\begin{aligned}\frac{d}{dt}SR_{W,k} = \; & \omega_G \sigma_G SI_{W,k} - \mu SR_{W,k} + \sigma_E(1-\omega_E)IR_{W,k} \\ & -c_k SR_{W,k}\left(\left(\pi_E^E + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^E + \pi_{EG}^{EG}\right) I_M^{EG}\right)\end{aligned} \tag{A.5}$$

$$\begin{aligned}\frac{d}{dt}RI_{W,k} = \; & -\sigma_G RI_{W,k} - \mu RI_{W,k} + \sigma_E \omega_E II_{W,k} \\ & +c_k RS_{W,k}\left(\left(\pi_E^G + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^G + \pi_{EG}^{EG}\right) I_M^{EG}\right)\end{aligned} \tag{A.6}$$

$$\begin{aligned}\frac{d}{dt}RS_{W,k} = \; & \omega_E \sigma_E IS_{W,k} - \mu RS_{W,k} + \sigma_G(1-\omega_G)RI_{W,k} \\ & -c_k RS_{W,k}\left(\left(\pi_E^G + \pi_E^{EG}\right) I_M^E + \left(\pi_G^G + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^G + \pi_{EG}^{EG}\right) I_M^{EG}\right)\end{aligned} \tag{A.7}$$

$$\frac{d}{dt}RR_{W,k} = \omega_G \sigma_G RI_{W,k} + \omega_E \sigma_E IR_{W,k} - \mu RR_{W,k} \tag{A.8}$$

$$SS_{W,k} + SI_{W,k} + IS_{W,k} + II_{W,k} + IR_{W,k} + SR_{W,k} + RI_{W,k} + RS_{W,k} + RR_{W,k} = 1 \tag{A.9}$$

where,

- R,S,I denote the proportion of individuals resistant/immune, susceptible and infected respectively.
- The first position used in the notation $XX_{W,k}$ represents the extragenital site and the second position the genital site. Hence, $RS_{W,k}$ represents the proportion of women immune at the extragenital site and susceptible at the genital site.
- W denotes women gender and M denotes men gender.
- The index k refer to the level/class of sexual activity (i.e., low vs. high).
- c_k is the effective rate of new partner acquisition in class k.
- ω_j is the probability of developing immunity upon clearing infection at site j.
- σ_j is the clearance rate (1/year) of infection at site j.
- λ_j is the rate of autoinoculation from site j to the other site.
- μ is the mortality rate (1/year).

Probabilities of transmission

- $\pi_j^{j'}$ is the probability that a man infected at site j infects a woman infection-free at the site(s) j' within a sexual partnership. Hence, we have:
  1. π_G^G: probability that an individual infected at the genital site transmit an infection at the genital site.
  2. π_G^E: probability that an individual infected at the genital site transmit an infection at the extragenital site.
  3. π_G^{EG}: probability that an individual infected at the genital site transmit an infection at both sites.
  4. π_E^G: probability that an individual infected at the extragenital site transmit an infection at the genital site.
  5. π_E^E: probability that an individual infected at the extragenital site transmit an infection at the extragenital site.
  6. π_E^{EG}: probability that an individual infected at the extragenital site transmit an infection at both sites.

7. π_{EG}^{G}: probability that an individual infected at both sites transmit an infection at the genital site.

8. π_{EG}^{E}: probability that an individual infected at both sites transmit an infection at the extragenital site.

9. π_{EG}^{EG}: probability that an individual infected at both sites transmit an infection at the both sites.

- If we denote $\beta_j^{j'}$ the probability per partnership of transmitting infection at site $j = \{\text{genital, extragenital}\}$ to site j', then we have:

$$\pi_G^G = \beta_G^G \left(1 - \beta_G^E\right) \tag{A.10}$$

$$\pi_G^{EG} = \beta_G^G \beta_G^E \tag{A.11}$$

$$\pi_{EG}^E = \left(1 - \left(1 - \beta_G^E\right)\left(1 - \beta_E^E\right)\right)\left(1 - \beta_G^G\right)\left(1 - \beta_E^G\right) \tag{A.12}$$

$$\pi_{EG}^{EG} = \left(1 - \left(1 - \beta_G^E\right)\left(1 - \beta_E^E\right)\right)\left(1 - \left(1 - \beta_G^G\right)\left(1 - \beta_G^E\right)\right) \tag{A.13}$$

The four parameters $\beta_j^{j'} : \{\beta_G^G, \beta_G^E, \beta_E^G, \beta_E^E\}$ are the one being calibrated plus the four other parameters for women-to-men transmission.

Rate of new partner acquisition

- c_k, the effective annual rate of new partner acquisition in class k, was computed using data from the National Survey of Family Growth 2011-2013 (NSFG)[272] on lifetime number of partners for both females and males aged 15-30. Importantly, the rate c_k does not represent the mean annual rate of new partner acquisition among the 15-30 population of the NSFG survey. In fact, an adjustment was made to compensate for the lack of heterogeneity in sexual activity in our model by using the computed variance in the number of lifetime partners. Formally,

$$c_k = r_k + \frac{\sigma_k}{r_k} \tag{A.14}$$

where r_k and σ_k are respectively the mean and variance of the annual rate of acquisition of new partners in sexual activity class k. The justification for using such c_k is that the R_0 of a population with random mixing and sexual activity parameters r_k and σ_k is equal to the R_0 of a population with random mixing and sexual activity parameters equal c_k and a zero variance (homogeneous population).

Mixing across sexual activity classes

- I_{gender}^j is the total proportion of infected at site j across sexual activity classes weighted by c_k:

$$\textit{(i.e., } I_W^j = w_1\left(RI_{W,1} + SI_{W,1}\right) + w_2\left(RI_{W,2} + SI_{W,2}\right)\textit{)} \tag{A.15}$$

$$\textit{(i.e., } I_M^j = w_1\left(RI_{M,1} + SI_{M,1}\right) + w_2\left(RI_{M,2} + SI_{M,2}\right)\textit{)} \tag{A.16}$$

where,

$$w_1 = \frac{N_1 c_1}{N_1 c_1 + N_2 c_2} \quad \text{and} \quad w_2 = \frac{N_2 c_2}{N_1 c_1 + N_2 c_2} \tag{A.17}$$

with N_k the number of individuals in class k. This is the standard formula of proportional mixing.

A.1.3 Model equations for systemic immunity

The set of differential equations A.18-A.24 is used to investigate the effect of multi-site transmission on vaccination impact assuming systemic immunity after genital infection only (scenario 3) or systemic immunity after genital and extragenital infection (scenario 4).

$$\begin{aligned}\tfrac{d}{dt}SI_{W,k} = \; & -\sigma_G SI_{W,k} - \mu SI_{W,k} + \sigma_E(1-\omega_E)II_{W,k} + c_k SS_{W,k}\left(\pi_E^G I_M^E + \pi_G^G I_M^G + \pi_{EG}^G I_M^{EG}\right) \\ & -c_k SI_{W,k}\left(\left(\pi_E^E + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^E + \pi_{EG}^{EG}\right) I_M^{EG}\right) - \lambda_G SI_{W,k}\end{aligned} \tag{A.18}$$

$$\begin{aligned}\tfrac{d}{dt}IS_{W,k} = \; & -\sigma_E IS_{W,k} - \mu IS_{W,k} + \sigma_G(1-\omega_G)II_{W,k} + c_k SS_{W,k}\left(\pi_E^E I_M^E + \pi_G^E I_M^G + \pi_{EG}^G I_M^{EG}\right) \\ & -c_k IS_{W,k}\left(\left(\pi_E^G + \pi_E^{EG}\right) I_M^E + \left(\pi_G^G + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^G + \pi_{EG}^{EG}\right) I_M^{EG}\right) - \lambda_E IS_{W,k}\end{aligned} \tag{A.19}$$

$$\begin{aligned}\tfrac{d}{dt}II_{W,k} = \; & -\sigma_E II_{W,k} - \mu II_{W,k} - \sigma_G II_{W,k} + c_k SS_{W,k}\left(\pi_E^{EG} I_M^E + \pi_G^{EG} I_M^G + \pi_{EG}^{EG} I_M^{EG}\right) \\ & +c_k SI_{W,k}\left(\left(\pi_E^E + \pi_E^{EG}\right) I_M^E + \left(\pi_G^E + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^E + \pi_{EG}^{EG}\right) I_M^{EG}\right) + \lambda_E IS_{W,k} \\ & +c_k IS_{W,k}\left(\left(\pi_E^G + \pi_E^{EG}\right) I_M^E + \left(\pi_G^G + \pi_G^{EG}\right) I_M^G + \left(\pi_{EG}^G + \pi_{EG}^{EG}\right) I_M^{EG}\right) + \lambda_G SI_{W,k}\end{aligned} \tag{A.20}$$

$$\tfrac{d}{dt}IR_{W,k} = \; -\sigma_E IR_{W,k} - \mu IR_{W,k} + \sigma_G \omega_G II_{W,k} \tag{A.21}$$

$$\tfrac{d}{dt}RI_{W,k} = \; -\sigma_G RI_{W,k} - \mu RI_{W,k} + \sigma_E \omega_E II_{W,k} \tag{A.22}$$

$$\tfrac{d}{dt}RR_{W,k} = \omega_E \sigma_E IS_{W,k} + \omega_G \sigma_G SI_{W,k} - \mu RR_{W,k} + \sigma_G RI_{W,k} + \sigma_E IR_{W,k} \tag{A.23}$$

$$SS_{W,k} + SI_{W,k} + IS_{W,k} + II_{W,k} + IR_{W,k} + RI_{W,k} + RR_{W,k} = 1 \tag{A.24}$$

where the symbols are defined as the same as previously.

A.1.4 Supplementary figures

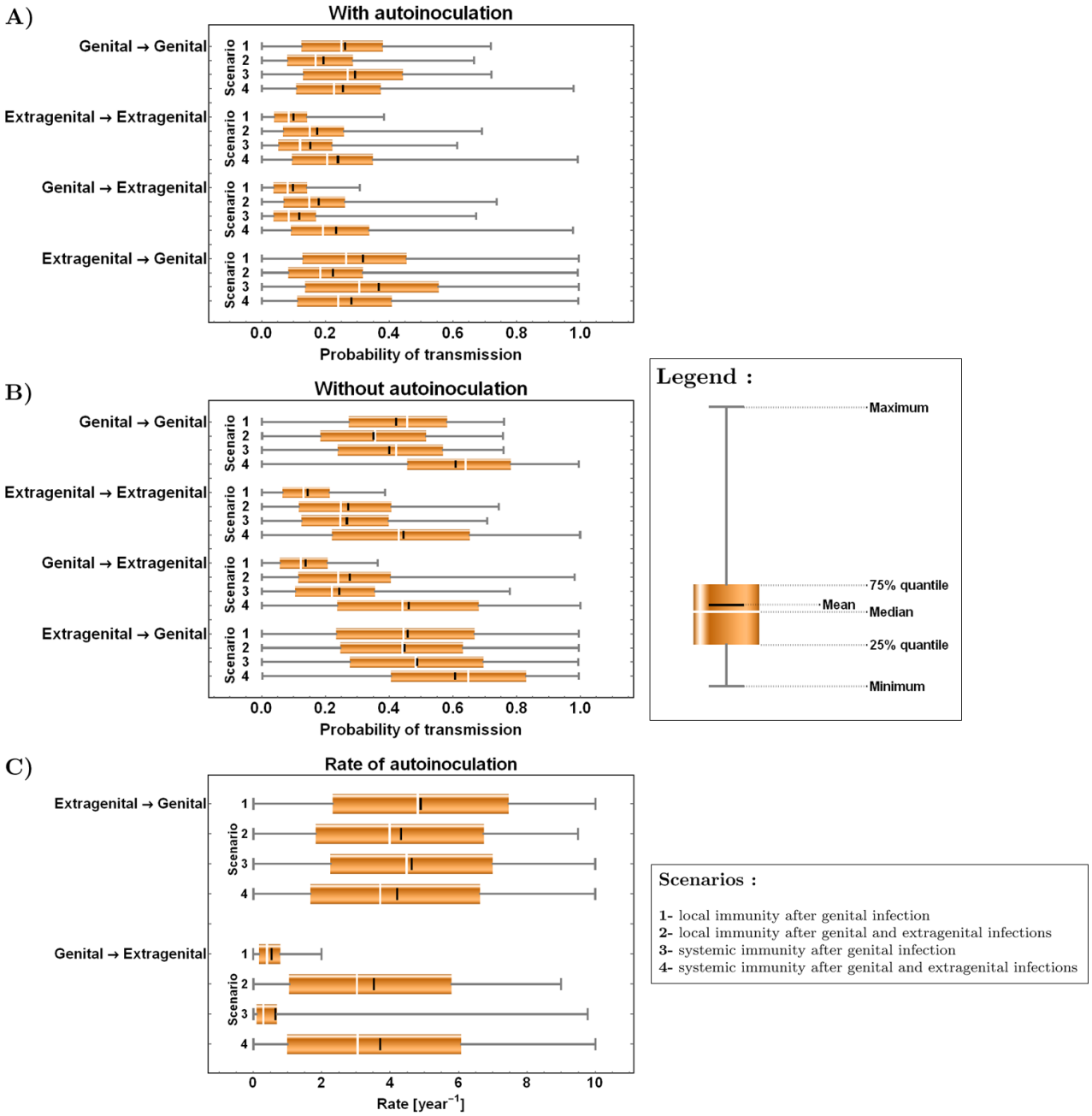


Figure A.5 – **Distributions of the probabilities of transmission and rates of autoinoculation identified in the calibration.** Uniform priors [0,1] were assumed for all probabilities of transmission and uniform priors [0,10] (year^{-1}) were assumed for the two rates of autoinoculation. Distributions for male-to-female probabilities of transmission and female-to-male probabilities of transmission were similar and thus were collapsed in one single distribution for each of the four probabilities of transmission.

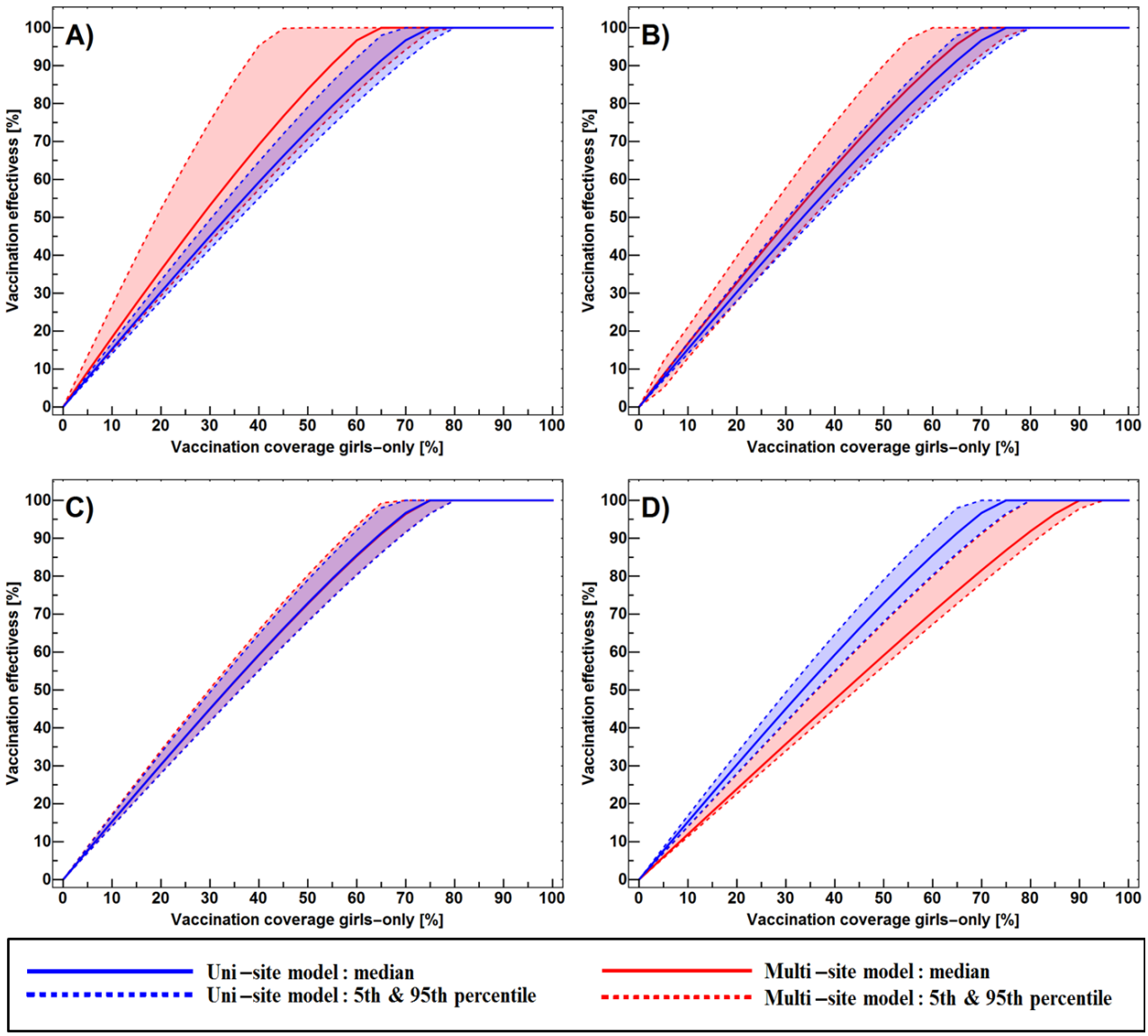


Figure A.6 – **Population-level effectiveness of HPV vaccination with a multi-site and uni-site models: comparison by vaccination coverage.** Vaccination program is girls-only and vaccine is assumed to have 100% efficacy and lifelong duration for the two sites of infection. No autoinoculation is assumed and four scenarios are considered regarding the development of natural immunity after infection: **A)** Local immunity after genital infection, **B)** Local immunity after genital and extragenital infection, **C)** Systemic immunity after genital infection, **D)** Systemic immunity after genital and extragenital infection. No autoinoculation is assumed. The reduction in HPV prevalence is computed at endemic equilibrium (i.e., 150 years) as: $\left(\text{Prev}_{\text{pre-vaccination}} - \text{Prev}_{\text{post-vaccination}}\right) / \text{Prev}_{\text{pre-vaccination}}$.

A.2 (Objective 2): Understanding the effect of the key factors of transmission responsible for divergences in predictions of HPV16 vaccination effectiveness between multi-site and uni-site models.

A.2.1 Models flow charts for the multi-site model

Figure A.7 presents the models flow chart (deaths/births and gender/sexual activity stratifications are not represented) used to conceptualize the simplified homogeneous multi-site model. The latter is used to explain the variance in predictions of vaccination effectiveness (Objective 2).

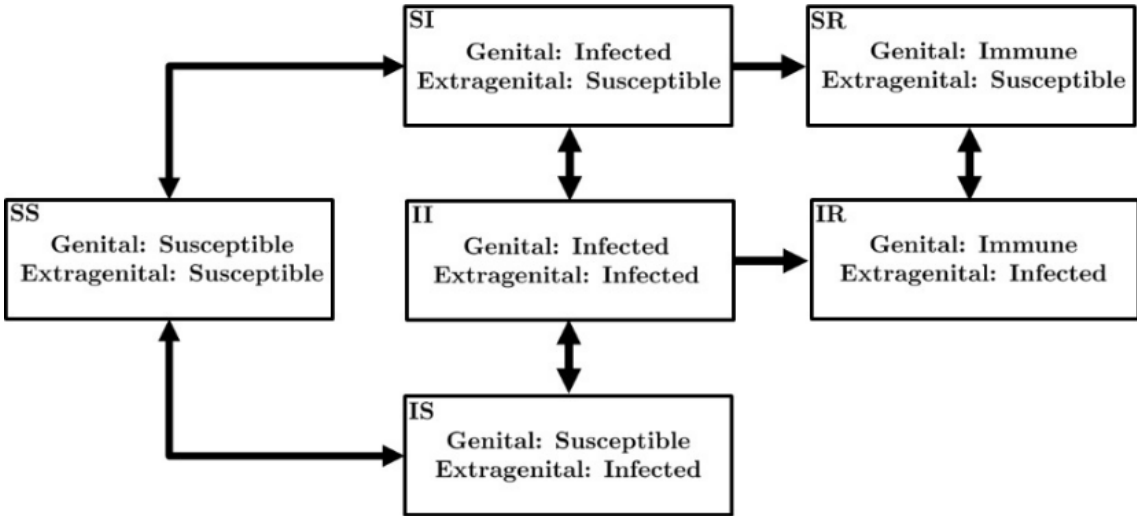


Figure A.7 – **Flow chart of multi-site model used to conceptualize the simplified homogeneous multi-site model.**

A.2.2 Model equations to assess the effect of transmission parameters on the reproductive number (R_0) of the multi-site model

The set of differential equations A.25-A.28 is used to investigate the effect of modifying transmission probabilities on the reproductive number R_0 with the multi-site model (see Figure 1.2 of the article).

$$\frac{\partial Y_1}{\partial t} = -\sigma Y_1 - \mu Y_1 + X_1 \left(\pi_{21} Y_2 + \pi_{11} Y_1\right) \tag{A.25}$$

$$\frac{\partial Y_2}{\partial t} = -\sigma Y_2 - \mu Y_2 + X_2 \left(\pi_{22} Y_2 + \pi_{12} Y_1\right) \tag{A.26}$$

$$\frac{\partial Z_1}{\partial t} = \omega\sigma Y_1 - \mu Z_1 \tag{A.27}$$

$$X_i = 1 - Y_i - Z_i \tag{A.28}$$

with,

- X_i the proportion susceptible at site $i = 1$ (genital), $i = 2$ (extragenital).
- Y_i is the proportion infected at site i.
- Z_1 is the proportion immune at site 1.
- $\pi_{i,j}$ is the effective contact rate for an infected at site i to a susceptible site j.
- σ is the clearance rate (1/year).
- ω is the probability of developing immunity upon clearance of infection at site 1.
- μ is the mortality rate (1/year).

Note that the proportion infected at two sites simultaneously can be omitted from these equations if we are only interested in the marginal proportions Y_1 and Y_2 (the sum of which can exceed 1). This key simplification arises from the additional assumption that an individual infected at two sites produces as many i infections as two individuals with a single infection at site 1 and 2. This is not true in general since the modes of transmission $1 \to i$ and $2 \to i$ are not necessarily independent.

A.2.3 R_0 and the effect of genital-extragenital/extragenital-genital transmission

The case of one gender (i.e., symmetrical between men and women)

The differential equations for the marginal proportions have the same structure as a classical model with sub-populations. Hence, existence and unicity of the system equilibrium has been established (see[273] for example).

At equilibrium, derivatives are equal to 0:

$$0 = -\sigma Y_1 - \mu Y_1 + X_1 \left(\pi_{21} Y_2 + \pi_{11} Y_1\right) \Longleftrightarrow \pi_{11} = \frac{\sigma + \mu}{X_1} - \frac{\pi_{21} Y_2}{Y_1}$$

$$R_{11} := \frac{\pi_{11}}{\sigma + \mu} = \frac{1}{X_1} - \frac{R_{21} Y_2}{Y_1} \tag{A.29}$$

$$0 = -\sigma Y_2 - \mu Y_2 + X_2 \left(\pi_{22} Y_2 + \pi_{12} Y_1\right) \Longleftrightarrow \pi_{22} = \frac{\sigma + \mu}{X_2} - \frac{\pi_{12} Y_1}{Y_2}$$

$$R_{22} := \frac{\pi_{22}}{\sigma + \mu} = \frac{1}{X_2} - \frac{R_{12} Y_1}{Y_2} \tag{A.30}$$

$$R_0 = \frac{R_{11} + R_{22}}{2} + \frac{1}{2}\sqrt{\left(R_{11} + R_{22}\right)^2 + 4\left(R_{12} R_{21} - R_{11} R_{22}\right)} \tag{A.31}$$

The formula for the reproductive number is derived from the definition: the dominant eigenvalue of the Next-Generation-Matrix[274]. R_{ij} are the number of secondary j infections caused by someone infected at i in a completely susceptible population. Hence, genital→extragenital transmission is captured by R_{12} and extragenital→genital transmission is captured by R_{21}. Furthermore, for this model, $\left(1 - \frac{1}{R_0}\right)$is the threshold above which homogeneous vaccination causes elimination of the infection in the population.

We consider the partial derivatives of R_0 as a function of both R_{12} and R_{21} assuming Y_1 and Y_2 are known at equilibrium.

$$\frac{\partial R_0}{\partial R_{12}} = -\frac{Y_1}{2Y_2} + \frac{-2\frac{Y_1}{Y_2}\left(R_{22}-R_{11}\right)+4R_{21}}{4\sqrt{\left(R_{11}+R_{22}\right)^2+4\left(R_{12}R_{21}-R_{11}R_{22}\right)}} \tag{A.32}$$

Hence,

$$\frac{\partial R_0}{\partial R_{12}} \geq 0 \Longleftrightarrow \frac{Y_1}{2Y_2} 4\sqrt{\left(R_{11}+R_{22}\right)^2+4\left(R_{12}R_{21}-R_{11}R_{22}\right)} \leq -2\frac{Y_1}{Y_2}\left(R_{22}-R_{11}\right)+4R_{21} \tag{A.33}$$

$$* \Longleftrightarrow R_{21}\left[\left(\frac{Y_1}{Y_2}\right)^2 R_{12} - R_{21} + \left(R_{22}-R_{11}\right)\frac{Y_1}{Y_2}\right] \leq 0 \Longleftarrow X_1 \leq X_2 \tag{A.34}$$

where we squared both side of the equations and replaced R_{11} and R_{22} as functions of R_{12} or R_{21}.

* There is theoretically the possibility that $\left(-2\frac{Y_1}{Y_2}\left(R_{22}-R_{11}\right)+4R_{21}\right) \leq 0$. By replacing R_{22} and R_{11} we see that this can never happen if $X_1 \leq X_2$.

We can similarly (symmetric) show that $X_1 \leq X_2 \Longrightarrow \frac{\partial R_0}{\partial R_{12}} \leq 0$.

Thus, if the number of susceptibles at equilibrium is lower for genital infections than extragenital infections, increasing genital→extragenital transmission increases R_0 while increasing extragenital→genital transmission decreases R_0 (for fixed Y_1, Y_2, X_1, X_2). In particular, if there is no natural immunity or if the number of immune individuals is the same for genital and extragenital infections, the condition on the number of susceptibles is equivalent to having a greater prevalence of genital infections compared to extragenital infections.

As a corollary, the minimum R_0 for given equilibrium $X_2 \geq X_1$ is obtained when genital→extragenital is minimal and extragenital→genital is maximal. Note that we did not need the fact that the clearance rates were the same for both sites in the proof. The value of R_0 for this multi-site model is comprised between $1/X_2$ and $1/X_1$. In fact, assuming that at equilibrium the matrix:

$$\begin{pmatrix} R_{11}X_1 & R_{21}X_1 \\ R_{12}X_2 & R_{22}X_2 \end{pmatrix} \tag{A.35}$$

is irreducible and $X_2 \geq X_1 > 0$ (without loss of generality), we have:

$$\begin{array}{rl} Y_1 = & R_{11}Y_1X_1 + R_{21}Y_2X_1 \\ Y_2 = & R_{22}Y_2X_2 + R_{12}Y_1X_2 \end{array} \Longleftrightarrow \begin{pmatrix} R_{11}X_1 & R_{21}X_1 \\ R_{12}X_2 & R_{22}X_2 \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \tag{A.36}$$

$$\Longrightarrow \rho\begin{pmatrix} R_{11}X_1 & R_{21}X_1 \\ R_{12}X_2 & R_{22}X_2 \end{pmatrix} = 1 \tag{A.37}$$

Furthermore,

$$\Longrightarrow \begin{pmatrix} R_{11}X_1 & R_{21}X_1 \\ R_{12}X_1 & R_{22}X_1 \end{pmatrix} \leq \begin{pmatrix} R_{11}X_1 & R_{21}X_1 \\ R_{12}X_2 & R_{22}X_2 \end{pmatrix} \leq \begin{pmatrix} R_{11}X_2 & R_{21}X_2 \\ R_{12}X_2 & R_{22}X_2 \end{pmatrix} \tag{A.38}$$

$$\implies X_1\rho\begin{pmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{pmatrix} \leq 1 \leq X_2\rho\begin{pmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{pmatrix} \iff \frac{1}{X_2} \leq R_0 \leq \frac{1}{X_1} \tag{A.39}$$

with ρ denoting the largest real eigenvalue (or spectral radius) and the Perron-Frobenius theorem is used to justify the implications ($\implies$). Note that this holds for an arbitrary number of sites/dimensions. Again, we did not need the clearance rates to be the same for both sites. We see that the minimum $\left(R_0 = \frac{1}{X_2}\right)$ can always be attained through a Next-Generation matrix of the form:

$$\begin{pmatrix} 0 & R_{21} \\ 0 & R_{22} \end{pmatrix} \tag{A.40}$$

The maximum $\left(R_0 = \frac{1}{X_1}\right)$ can trivially be attained through the uni-site model:

$$\begin{pmatrix} R_{11} & 0 \\ 0 & 0 \end{pmatrix} \tag{A.41}$$

The case of two genders (i.e., asymmetrical between men and women)

We use the subscript $\{M, W\}$ to represent men and women respectively. Thus, we write Y_{iM} and Y_{iW} for the prevalence among men and women respectively with i representing the site of infection as previously. Then, we have the equations:

$$\frac{\partial Y_{1W}}{\partial t} = -\sigma Y_{1W} - \mu Y_{1W} + X_{1W}\left(\pi_{21M}Y_{2M} + \pi_{11M}Y_{1M}\right) \tag{A.42}$$

$$\frac{\partial Y_{2W}}{\partial t} = -\sigma Y_{2W} - \mu Y_{2W} + X_{2W}\left(\pi_{22M}Y_{2M} + \pi_{12M}Y_{1M}\right) \tag{A.43}$$

$$\frac{\partial Z_{1W}}{\partial t} = \omega\sigma Y_{1W} - \mu Z_{1W} \tag{A.44}$$

$$X_{iW} = 1 - Y_{iW} - Z_{iW} \tag{A.45}$$

with analogue equations for Y_{iM}.

For this two-gender model, we have the following bounds on R_0:

$$\frac{1}{\sqrt{X_{2M}X_{2W}}} \leq R_0 \leq \frac{1}{\sqrt{X_{1M}X_{1W}}} \tag{A.46}$$

This is similar to the one-gender case except that we take the geometrical mean of the proportion of susceptible men and women. In fact, we obtain the Next-Generation matrix:

$$\begin{pmatrix} 0 & 0 & R_{11M}X_{1W} & R_{21M}X_{1W} \\ 0 & 0 & R_{12M}X_{2W} & R_{22W}X_{2W} \\ R_{11W}X_{1M} & R_{21W}X_{1M} & 0 & 0 \\ R_{12W}X_{2M} & R_{22W}X_{2M} & 0 & 0 \end{pmatrix} \tag{A.47}$$

Then we proceed as previously using the fact that the eigenvalues of the matrix composed of blocks $A = 0$, λB, γC, $D = 0$ are:

$$\text{eigenvalues}\begin{pmatrix} 0 & \lambda B \\ \gamma C & 0 \end{pmatrix} = \text{eigenvalues}\begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix}\sqrt{\lambda\gamma} \tag{A.48}$$

As for the effect of R_{12G} and R_{21G} ($G = \{M, W\}$), if the prevalence and proportion of susceptibles are symmetric, i.e. $Y_{1M} = Y_{1W}$, $X_{1M} = X_{1W}$, $Y_{2M} = Y_{2W}$, $X_{2M} = X_{2W}$, then the same results holds as in the one-gender case, even if the transmission parameters are asymmetric. Hence, if $X_{1G} \leq X_{2G}$, then

$$\frac{\partial R_0}{\partial R_{12G}} \geq 0 \text{ and } \frac{\partial R_0}{\partial R_{21G}} \leq 0 \tag{A.49}$$

the one-gender case being a particular case. Computing $\left(\frac{\partial R_0}{\partial R_{21M}}\right)$ we get:

$$\begin{aligned}\frac{\partial R_0}{\partial R_{21M}} = \quad & R_{12W} - \frac{R_{11W}Y_{2M}}{2Y_{1M}} + \frac{-2\left(\frac{Y_{2M}}{Y_{1M}X_{2M}}\right)(R_{12W}R_{21W} - R_{22W}R_{11W}) + 2\left(R_{12W} - \frac{R_{11W}Y_{2M}}{2Y_{1M}}\right)(R_{12W}R_{21M} + \ldots)}{\sqrt{-4(R_{12W}R_{21W} - R_{22W}R_{21M})(R_{12M}R_{21M} - R_{22M}R_{11M}) + \ldots}} \\ & \frac{(\ldots + R_{12M}R_{21W} + R_{11W}R_{11M} + R_{22W}R_{22M})}{\ldots + (R_{12W}R_{21M} + R_{12M}R_{21W} + R_{11W}R_{11M} + R_{22W}R_{22M})^2}\end{aligned} \tag{A.50}$$

We proceed as before:

$$\frac{\partial R_0}{\partial R_{21M}} \leq 0 \Longleftrightarrow \left(-R_{12W} + \frac{R_{11W}Y_{2M}}{2Y_{1M}}\right) \leq \ldots \tag{A.51}$$

To take the square of both sides, it suffices to show that if

$$R_{12W} - \frac{R_{11W}Y_{2M}}{2Y_{1M}} \geq 0 \tag{A.52}$$

then the numerator of the second term is negative. This can be shown by replacing R_{12W} by $\frac{R_{11W}Y_{2M}}{2Y_{1M}}$ using equation A.52 and $R_{12W}R_{21M} + R_{12M}R_{21W} + R_{11W}R_{11M} + R_{22W}R_{22M}$ by $\frac{2}{X_{1M}^2}$ using the formula for R_0. After taking the square of both sides, we get:

$$\begin{aligned}& 16\left(X_{1M} - X_{2M}\right)\left(R_{12W}X_{2M}Y_{1M}^2 + Y_{2M}\left(R_{21W}X_{1M}Y_{2M}\right)\right)\ldots \\ & \left(R_{12W}X_{1M}Y_{1M}\left(R_{12M}X_{2M}Y_{1M} - Y_{2M}\right) + R_{12W}X_{2M}Y_{2M}\left(R_{21W}X_{1M}Y_{2M} - Y_{1M}\right)\right)\ldots \\ & \left(R_{12W}X_{1M}X_{2M}Y_{1M}^2 + Y_{2M}\left(-\left(X_{1M} + X_{2M}\right)Y_{1M}\right) + R_{21W}X_{2M}X_{1M}Y_{2M}\right) \leq 0 \\ & \Longrightarrow \frac{\partial R_0}{\partial R_{21M}} \leq 0\end{aligned} \tag{A.53}$$

The first term of the product is negative by hypothesis, the second is positive and the last two terms are negative as easily seen. The procedure is similar for $\frac{\partial R_0}{\partial R_{12M}}$. Importantly, equation A.49 needs not be true if prevalences are asymmetric. As a counterexample, consider:

$$\begin{aligned} X_{1M} &= 0.5 \\ X_{2M} &= 0.6 \\ X_{1W} &= 0.2 \\ X_{2W} &= 0.5 \\ Y_{iG} &= 1 - X_{iG} \\ R_{12W} &= 0.6 \\ R_{21W} &= 2 \\ R_{12M} &= 1 \\ R_{21M} &= 10 \end{aligned} \tag{A.54}$$

One can check that $\frac{\partial R_0}{\partial R_{12M}} \leq 0$. However, our numerical simulations with asymmetrical prevalences of HPV:

$$\begin{array}{ll} Y_{1M} = & [0.05 - 0.07] \\ Y_{2M} = & [0.01 - 0.03] \\ Y_{1W} = & [0.05 - 0.07] \\ Y_{2W} = & [0.01 - 0.03] \end{array} \tag{A.55}$$

did not produce significant violations of equation A.49. Hence, equation A.49 appears robust to small deviation of symmetry.

A.2.4 Model with simultaneous transmission: violation of the upper bound on R_0.

Assume that equilibrium prevalences at n sites are given by $Y_1, \ldots, Y_n$ and that for the sake of the example there is no natural immunity and thus the proportion of susceptible is given by $1-Y_1, \ldots, 1-Y_n$. According to the results presented in the section 5, we should have that the upper bound on R_0 (as defined by 1-1/epidemic threshold) is:

$$\frac{1}{1 - \max_i(Y_i)} \tag{A.56}$$

We show by an example that R_0 can be arbitrary high even though $\max_i(Y_i) \leq k$ for any k, so that the result in section 5 is false if there is simultaneous transmission (or autoinoculation).

Let c be the contact rate, D_i the average duration of infection at sites $i = 1, \ldots, n$. Assume that the probability of transmission is 100% per contact between any two sites. Hence, someone who becomes infected is always infected at all sites simultaneously. For any given R_0 and any maximum prevalence $k = \max_i(Y_i)$, pick $D_i = D$ so that:

$$k \geq cD \tag{A.57}$$

c is also the maximum force of infection at site i if everyone in the population is infected at any site $(\text{proportion infected} \cdot \text{contact rate} \cdot \text{probability of transmission}) = (100\% \cdot c \cdot 100\%) = c$.

Thus, at equilibrium, the prevalence at site i will be lower or equal to k if the equation A.57 is satisfied. Pick n the number of sites, so that:

$$\sum_{i=1}^{n} \frac{D}{i} \geq \frac{R_0}{c} \tag{A.58}$$

Note that $\sum_{i=1}^{n} \frac{D}{i}$ is the average duration for someone infected at all sites to clear all his infections. The number of co-infections within an individual has no importance on transmission since the probability per contact is 100% for any pair of sites. Hence, an individual has the same infectious potential as long as one site is infected. We thus see that the reproductive number for this model is:

$$c \sum_{i=1}^{n} \frac{D}{i} \tag{A.59}$$

which is greater than the given R_0. If we add mortality, the maximum R_0 would be bounded by c times the life expectancy.

A.2.5 Supplementary figures

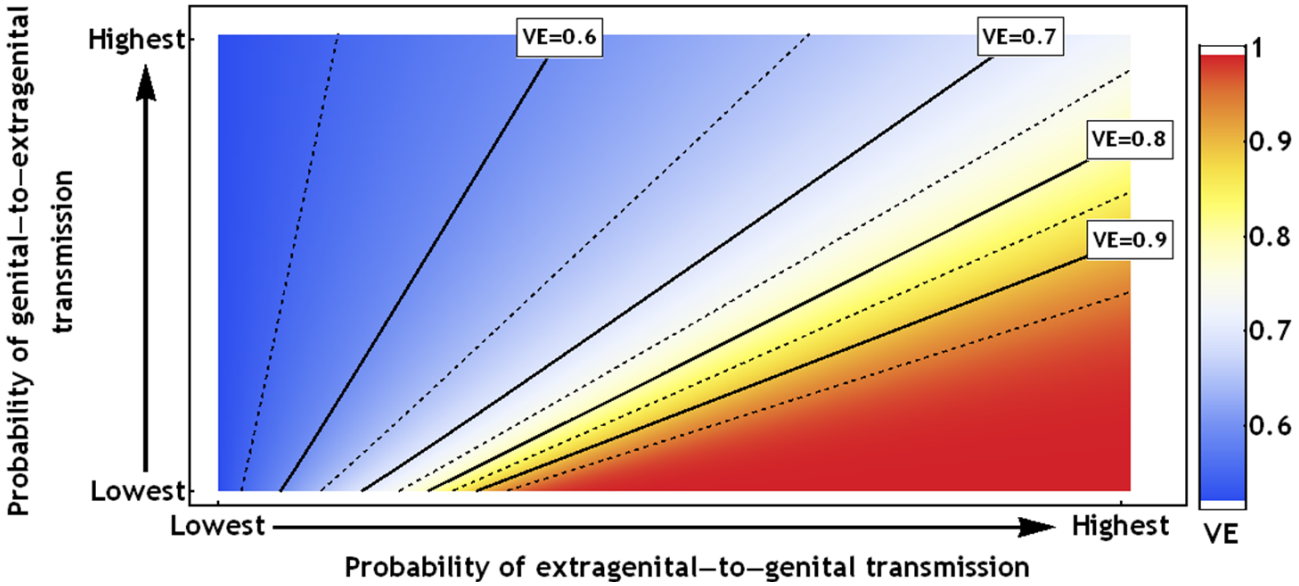


Figure A.8 – **Effect of transmission parameters on predicted vaccination effectiveness of the multi-site model.** Model simplifications (i.e., no simultaneous transmission of both extragenital and genital infections, homogeneous sexual activity, symmetric with respect to gender) were made to enable computation of all the items using analytic formulae. Model was calibrated to a genital prevalence of 7% and an extragenital prevalence of 3%. The x-axis represents probability that a male with an extragenital infection infects the genital site of his partner during a partnership, and the y-axis represents the probability that a male with a genital infection infects the extragenital site of his partner during a partnership. (VE : vaccination effectiveness computed as: $VE = \frac{\text{Prevalence genital site pre vaccination-Prevalence genital site post vaccination}}{\text{Prevalence genital site pre vaccination}}$)

Appendix B

Technical appendix. Impact of changes in sexual behavior on past and future trends of HPV infections and related cancers: Model description

B.1 Overview

We aimed to simulate the sexual behavior and HPV transmission between the ages of 12 and 44 years of white men and women from the US population born between 1850 and 1999. To do this, we considered a fictive population of women and men born every year from 1850 to 1999. Sexual behavior and HPV transmission was reproduced between the ages of 12 and 44 years. In 2015, the follow-up relevant to sexual behavior and HPV transmission ends for all individuals remaining. No mortality is assumed between 12-44 years old. The time-step for the simulation was a month as the temporal resolution of the data was also monthly. All individuals are identified by ids. Each of the individuals has the following dates recorded:

Date of birth
Dates of every start of cohabitation with a sexual partner
Dates of every end of cohabitation with a sexual partner
Dates of every non-cohabiting sexual partnership

Hence, each individual has a history of cohabitations and non-cohabiting sexual partnerships. Furthermore, every sexual partner of an individual is identified by its id if that partner meets the requirements to be in the model (i.e., aged between 12-44 years old and born between 1850 and 1999). If not, the age, the race, and the level of sexual activity of the sexual partner are recorded. Importantly, we proceeded differently to estimate the sexual history of those born before 1944 than for those born after 1945. This is because the main surveys we used, the National Survey of Family Growth (NSFG) and the National Health and Social Life Survey (NHSLS), do not provide data on birth cohorts before 1944. Hence, to inform the sexual activity of cohorts born before 1944, we used data from fewer vari-

ables from older individuals (aged > 50 years old) from the General Social Survey (GSS). Additional assumptions were needed for these older birth cohorts.

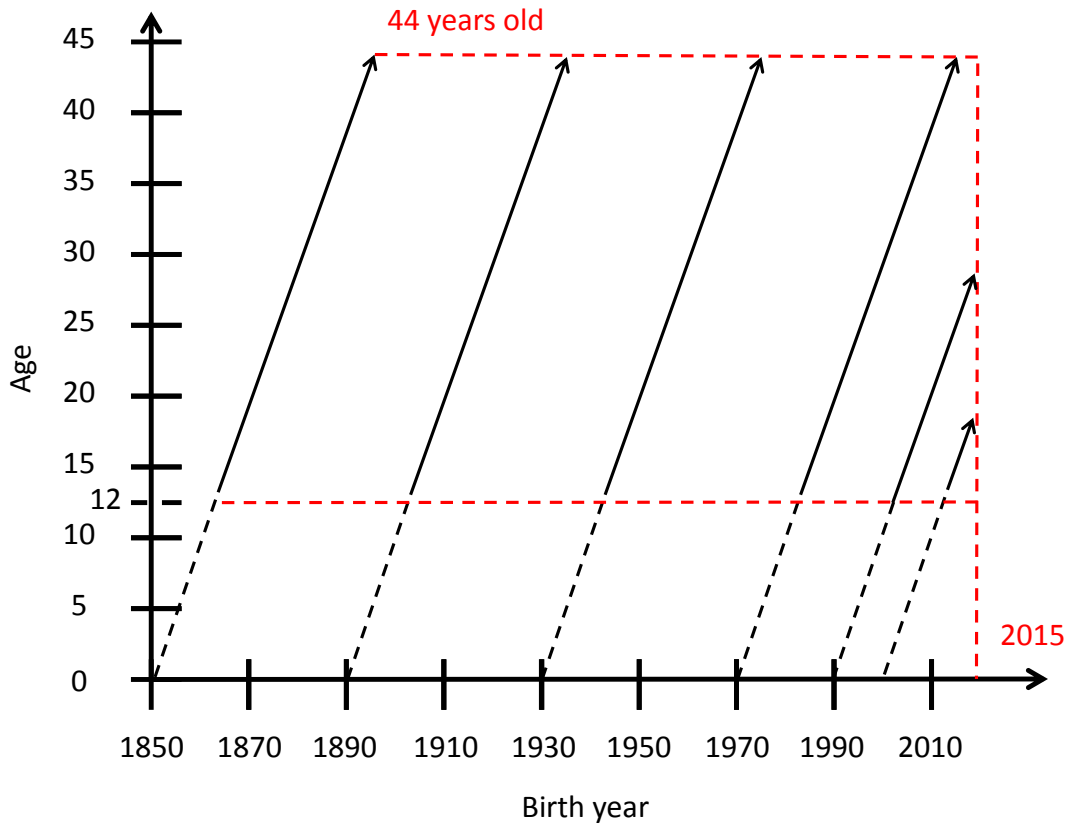


Figure B.1 – **Age-cohort illustration of the model population**. The earliest birth cohort of the model was born in 1850, the latest in 1999. Individuals are simulated until they reach 44 years of age or until 2015, whichever comes first.

B.1.1 Demographics

From 1850 to 1999, for each race (Non-Hispanic White, Non-Hispanic Black, Hispanic and Other), each year of age, each calendar year and each gender, we determined the size of the population. To do this, we used data from the US decennial census[224] to estimate the population sizes for 21 different points: 1) the 0-4 years old population for every decade from 1900 to 2000, 2) the 35-44 years old population for every decade from 1940 to 2010, 3) the 25-34 years old population in 2015, and 4) the 15-24 years old population in 2015. Then, we used linear interpolation to estimate the size of the population for the remaining points. Finally, we assumed that the birth cohorts between 1850 and 1900 were all identical, i.e., the population is stable.

Figure B.2 – **Size of the white male and female population in the US by birth year**.

B.1.2 Sexual behavior surveys

The surveys and the information used to inform the model on the sexual activity of the US population are presented in Table B.1.

Table B.1 – **Selected national surveys on the sexual behavior of the US population**

Surveys	**Surveys years**[a]	**Number of participants**	**Age of participants**	**Birth cohorts covered**	**Selected sexual behavior variables**	**Interview method**
General Social Survey	1988-91, 93 94, 96, 98, 00 02, 04, 06, 08	$\approx$ 30000	$>$18 yrs	1900-1999	1. Number of sexual partners after 18 years old 2. Age at first marriage	Self-Administered (Computer & pen & paper)
National Health and Social Life Survey	1991	$\approx$ 2000	18-59 yrs	1931-1973	1. Dates of cohabitation with sexual partners 2. Age of cohabitation partners 3. Age at first sexual intercourse 4. Lifetime number of sexual partners 5. Race of sexual partners 6. Timing of sexual partnerships relative to dates of cohabitation (e.g., before the first cohabitation	Self-Administered (pen & paper)
National Survey of Family Growth	1988, 95, 2002, 2006-10, 2011-13, 2013-15	$\approx$ 50000	15-44 yrs	1945-1999	1. Dates of cohabitation with sexual partners 2. Age and race of cohabitation partners 3. Age at first sexual intercourse 4. Age and race of first sexual intercourse partner 5. Lifetime number of sexual intercourse partners 6. Past year number of sexual intercourse partners 7. Age of past year sexual intercourse partners 8. Ever performed oral sex[b]	ACASI[e] & CAPI[f]
Three surveys from France	1972, 1993, 2008	$\approx$ 19000	18-69[c] yrs	1903-1990	1. Ever received cunnilingus 2. Ever received fellatio	Telephone interview & Face-to-face interview

[a]Only years of surveys with sexual behavior data are indicated; [b]Years 2006-2010, 2011-2013 and 2013-2015 only; [c]Some questions for some years had respondents aged 60-69 years; [d]Since the 2009-2010 editions; [e] Audio Computer-Assisted Self-Interview; [f]Computer-Assisted Personal Interviewing;

B.2 Women's sexual history for birth cohorts born between 1945 and 1999

B.2.1 Overview

We estimated all the events dates in the sexual history of women in the following order: 1) date of first sex, 2) dates of start and end of cohabitations in chronological order, 3) dates of non-cohabiting sexual partners. The order is important because estimation of the dates of any event has to be conditional on dates already estimated. For instance, the date of the start of the first cohabitation is conditional on the date of first sex. Likewise, estimation of 3) dates of non-cohabiting sexual partners is conditional on 1) date of first sex and 2) dates of cohabitations. It is crucial to preserve the dependency structure between the different events of the sexual history to accurately reproduce sexual history profiles.

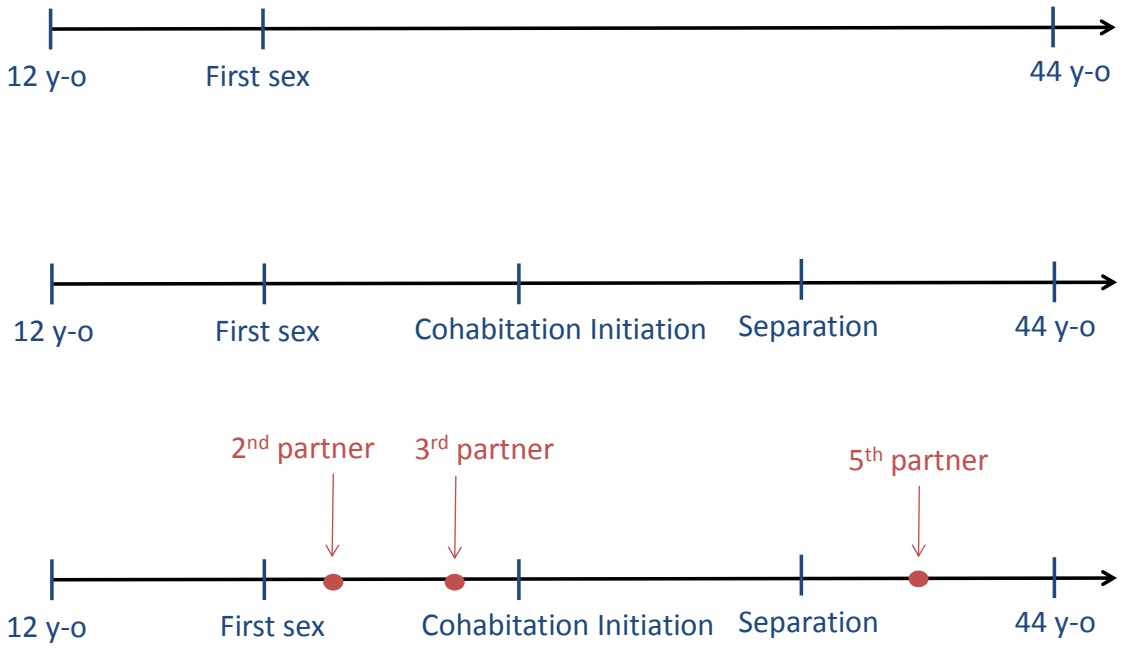


Figure B.3 – **Illustration of the building of a sexual history profile of a woman**. The woman in this example had 5 lifetime sexual partners: 4 non-cohabiting partners and 1 cohabitation partner. The date of first sex is estimated first, then dates of cohabitations (possibly more than one), then dates of non-cohabiting sexual partnerships (red). Non-cohabiting sexual partnerships are modelled as instantaneous.

B.2.2 Date of first sex

The NSFG surveys of 1982, 1988, 1995, 2002, 2006, 2011-2013, 2013-2015 provide data on the age at first sex of women born between 1939 and 1999. Using these data, we fitted a statistical model of the rate of sexual initiation per age and birth year. The model was fitted using "gamlss" function of the "gamlss.cens" package. The models specifications as well as illustrations of the fit are presented in the Technical Appendix C.1.1.

B.2.3 Dates of cohabitation

The NSFG surveys of 1988, 1995, 2002, 2006, 2011-2013, 2013-2015 provide data on the cohabitation history of women born between 1945 and 1999. Dates of cohabitations were recorded for all women up to a maximum varying between 6 and 10 depending on the survey. Using these data, dates of cohabitation were estimated in chronological order, starting with the date of the start of the first cohabitation and ending with the date of the end of the last cohabitation. A maximum number of 8 cohabitations was assumed because the proportion of women with more than 8 was negligible. Statistical models of the rates of cohabitation initiation and separation were fit to data using as predictors the dates already estimated. These models are described in Table B.2 with complete formulation and illustrations of fits given in the Technical Appendix C.1.2.

B.2.4 Dates of non-cohabiting sexual partnerships

Estimating the number of non-cohabiting sexual partnerships

We first estimated the number of non-cohabiting sexual partnerships given the history of cohabitation and the age of first sex of each woman. The NSFG surveys of 1988, 1995, 2002, 2006-2010, 2011-2013, 2013-2015 provide data on the distribution of the number of non-cohabiting sexual partnerships. Statistical models were fit to the number of non-cohabiting sexual partnerships using as predictors a characterization of cohabitation history and age at first sex. To characterize cohabitation history, we used the following variables:

Age at first sex
The number of months between first sex and start of first cohabitation (the pre-cohabitation period)
The number of months outside a cohabitation partnership after the 1st cohabitation (the post-cohabitation period)
The number of cohabitations
Age

Months outside a cohabitation partnership are at greater "risk" for the occurrence of non-cohabiting partners: the number of such months is a very strong predictor of the total number of non-cohabiting partners. The length of pre-cohabitation is likewise a very strong predictor of the number of pre-marital partners. We considered two variants of the number of months outside a cohabitation partnership: the number of months unweighted, and the number of months weighted proportionally with respect to age so that younger ages are given more weight than older ages.

Sexually active white women were split into four classes according to their lifetime number of non-cohabiting partners: $0-2$, $3-10$, $11-40$, ≥ 41. We fitted four logistic regression models to determine class membership conditional on cohabitation history. We fitted two negative binomial regression models to determine the number of non-cohabiting partners within the $3-10$ and $11-40$

Table B.2 – **Predictors used to model time of initiation and separation of sexual partnerships involving cohabitation**

	Predictors
First cohabitation	
Initiation	Age at first sex, birth year, age
Separation	Age at first sex, birth year, age at 1st cohabitation, duration of cohabitation
Second cohabitation	
Initiation	Age at first sex, birth year, age at 1st cohabitation, age at end of 1st cohabitation, time since end of 1st cohabitation
Separation	Age at first sex, birth year, age at 2nd cohabitation, duration of 1st cohabitation, duration of current cohabitation
Third cohabitation or more	
Initiation	Age at first sex, birth year, age at previous cohabitation, age at end of previous cohabitation, number of previous cohabitations, time since end of previous cohabitation
Separation	Age at first sex, birth year, age at start of current cohabitation, duration of previous cohabitation, number of previous cohabitations, duration of current cohabitation

classes conditional on cohabitation history. For the open ≥ 41 class, we fitted a power law regression model using data from NSFG 95 and NHSLS which are not right-censored (i.e., there is no limit on the number of partners that can be reported). This process is illustrated in Figure B.4. Complete models formulations and illustrations of fits are given in the Technical Appendix C.1.3.

Determining the period of occurrence of non-cohabiting sexual partnerships

Given the previously estimated number of non-cohabiting partnerships and cohabitation history of each woman, we attributed each of the non-cohabiting partnership to one of 17 possible periods:

1. Pre-cohabitation: between date of first sex and date of start of first cohabitation
2. First cohabitation: between date of start of first cohabitation and date of end of first cohabitation
3. First post-cohabitation: between date of end of first cohabitation and the date of the start of second cohabitation

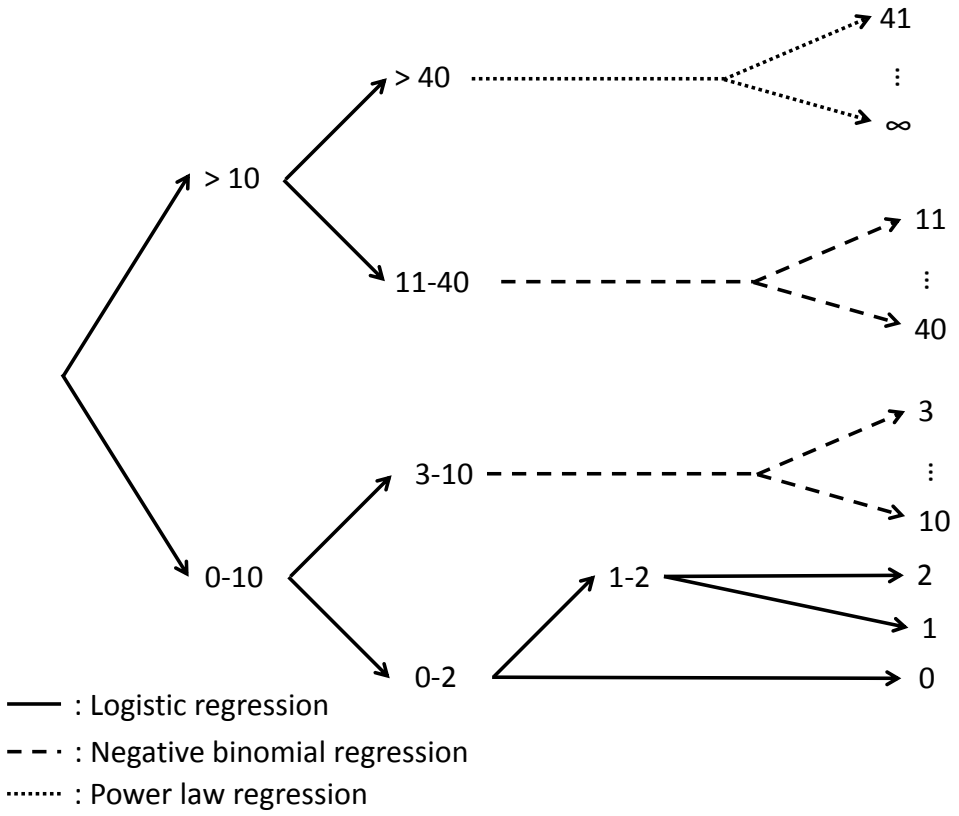


Figure B.4 – **Process of estimating the number of non-cohabiting sexual partnerships among women.** Women are split into four classes using logistic regression ($0-2$, $3-10$, $11-40$, ≥ 41). Women are first split into two classes using logistic regression ($1-10$, ≥ 11). Then each of these two classes is further split into two classes using logistic regression resulting in four final classes. Two negative binomial regressions are used to estimate the number of non-cohabiting sexual partners within the $3-10$ and $11-40$ classes. Using non-censored data from NSFG 95 and NHSLS, we also fitted a power law regression model for the open class ≥ 41.

⋮
16. Eighth cohabitation: between the date of start of eighth cohabitation and date of end of eighth cohabitation
17. Eighth post-cohabitation: between the date of end of eighth cohabitation and the end of follow-up

To do this, we used data from the NHSLS. In this nationally representative survey, complete cohabitation history is recorded as in the NSFG. Furthermore, lifetime sexual partners are classified according to the period in which they occurred. The six periods used in NHSLS to classify sexual partners are:

Pre-cohabitation
First cohabitation
First post-cohabitation
Second cohabitation
Second post-cohabitation: between the end of second cohabitation and one year before interview
Past year: up to one year before interview

We fitted to these data models:

$$P[t_i = k] = \frac{W_k}{\sum_j W_j}$$

where t_i is the period in which occurred the i^{th} non-cohabiting partner, and W_j is the weight of period j. Each weight W_j is given by a function:

$$log(\frac{W_j}{N_j}) = f(\beta_j, Predictors)$$

where N_j is the number of months in period j and $f(\beta_j, Predictors)$ is a piecewise-linear function with parameters to be estimated. The model described corresponds to acquiring non-cohabiting partners each month according to a Poisson distribution of mean proportional to

$$exp(f(\beta_j, Predictors))$$

White women in the NHSLS were divided in four classes according to their number of non-cohabiting partners: 1, $2-3$, $4-7$, ≥ 8. For each class, a model as described above was fitted by minimizing the log-likelihood using the fminsearch Matlab function. There are two key hypotheses to this model: 1) the timing of non-cohabiting partnership events has not changed between birth cohorts of 1900 to 1999 if we condition on the total number of lifetime partners and cohabitation history, 2) the t_i are independent one of another. In lay terms, 1) means that given two women with the same cohabitation history and the same lifetime number of partners but born at different times, both will have the same stochastic distribution governing the timing of non-cohabiting partners events. As for 2), it means that acquiring a partner during a given month does not lower the chance of acquiring a partner in the next month. In other words, we do not account for potential time-dependency between non-cohabiting partnerships (e.g., due to non-zero length of partnership). Complete models formulations and illustrations of fits are given in Technical Appendix C.1.3.

Determining the months of occurrence of non-cohabiting sexual partnerships occurring in a period outside a cohabitation partnership

After having classified sexual partnerships in one of the 17 periods described above, we determined the month during which every partnership occurred. For partnerships occurring in a period outside

a cohabitation partnership (i.e., odd periods), we attributed a weight to every year of age within the period and sampled the months of every partnership through a weighted random sampling with replacement.
To estimate these weights, we used data from the NSFG 95 on the date of sexual partnerships that occurred between the time of interview and 5 years before the time of interview. For every woman and every year of age available, we estimated the average number of new partners per month outside cohabitation during the year: $N_{i,year}$. To adjust for possible confounding effect of an association between the level of sexual activity and the years of age contributed (e.g., women with high number of partners having a lower age at first sex), we computed the ratio $RR_{i,year} = N_{i,year}/N_{i,year+1}$ for each woman and combined the ratios using Mantel-Haenszel to obtain an adjusted ratio RR_{year}. The weight of year k is thus the product $RR_0 \times \cdots \times RR_{k-1}$.

Determining the months of occurrence of non-cohabiting sexual partnerships occurring during a cohabitation partnership

Given a non-cohabiting partnership occurring during a given cohabitation partnership, we sampled the *t = month of occurrence of the non-cohabiting sexual partnership* given *l = the length of the cohabitation* from a conditional distribution. We used data from the NSFG 95 to estimate the joint distribution of t and l. We first identified all cohabitations that were formed and dissolved within the 5 year period before the interview date. We excluded ongoing cohabitations to prevent selection bias. We then identified all new sexual partnerships occurring during these cohabitations and we computed the timing of these new sexual partnerships, as the number of months since the start of cohabitation / length of cohabitation. We categorized l, the length of cohabitation, as short (< 1 year), medium ([1,2] years) and long (> 2 years and more). We used kernel density estimation with a Gaussian kernel to estimate the joint distribution of the timing of new sexual partnerships and the length of cohabitation.

B.3 Women's sexual mixing with men for birth cohorts born between 1945 and 1999

B.3.1 Overview

We determined the age and race of every woman's partner separately for cohabitating and non-cohabiting partners. Both the age and race of every woman's partner was dependent on the woman's sexual history. Complete models formulations and illustrations of fits are given in Technical Appendix C.2.1.

B.3.2 Cohabitation partners

Race of male partner

We used data on the race of each cohabitation partner in NHSLS, NSFG 2006-2010, 2011-2013, 2013-2015 to fit three logistic regression models. Race is categorized as Non-Hispanic White, Non-Hispanic Black, Hispanic and Others. The first model is used to estimate the probability that the partner is White. The second model estimates the probability that the partner is Black conditional on the partner not being White. The last model estimates the probability that the partner is Hispanic conditional on the partner being neither White nor Black. For all three models, predictors were the birth year, duration of cohabitation, lifetime number of partners at the time of interview, lifetime number of cohabitations at the time of interview, age at interview.

Age of male partner

We used data on the age of each cohabitation partner in NSFG 1988, 1995, 2002, 2006-2010, 2011-2013, 2013-2015 to fit four regression models of mixture of two normal law, one logistic regression model, and one regression on the mean of a gamma distribution. The logistic model estimates the odds that a woman has a male partner more than 25 years older than her. If the age difference is greater than 25 years, the gamma regression estimates the age difference between the male and female partner. If the age difference is under 25 years, one of the four models of mixture of two normal law estimates the age difference. These four models correspond to four categories of age at cohabitation: $12-19$, $20-26$, $27-34$, ≥ 35 years old. For all four models, predictors of the mean were the birth year, age at the time of cohabitation initiation, predictors of the logistic regression model of the mixture were duration of cohabitation, birth year, lifetime number of partners at the time of interview, lifetime number of cohabitations at the time of interview, age at interview.

B.3.3 Non-cohabiting partners

Complete models formulations and illustrations of fits are given in Technical Appendix C.2.2.

Race of male partner

We used data on the race of lifetime non-cohabiting partners only available in the NHSLS and demographic data on the race composition of the US population by birth cohort from decennial census to fit a model of preferential mixing. The model is used for extrapolation because the NHSLS data do not cover the younger birth cohorts. The formulation of the model is:

$$P_y[Race = k] = \frac{N_{k,y} W_k}{\sum_j N_{j,y} W_j}$$

with k in {White, Black, Hispanic, Others}, y is the birth year of the woman, $N_{k,y}$ is the population size of men of race k in birth cohort y, W_k is a preferential weight of white women towards men of race k, and $P_y[Race = k]$ is the probability that a new male partner of a woman born year y is of race k. The weights W_k are modelled as:

$$W_k = exp(\alpha_k + \beta_k log(LP))$$

with LP = Lifetime number of partners. Furthermore, we include a random individual effect to account for intra-individual correlation in race preference. We thus model the logit of the odds for a specific woman i as:

$$logit(P_y[Race = k]_i) = logit(P_y[Race = k]) + \epsilon_{i,k}$$

with $\epsilon_{i,k} \sim N(0,\alpha'_k)$ The model was fitted by minimizing the log-likelihood using the fminsearch Matlab function.

Age of male partner

We used data on the age of non-cohabiting partners available in NSFG 1995, 2002, 2006-2010, 2011-2013, 2013-2015. To avoid selection bias due to the length of partnership, we used only non-cohabiting partnerships which began during the retrospective follow-up period (either one-year or 5-years period before interview).

We proceeded in two steps. First, we fitted four regression models of mixture of two normal laws for the age difference between the male and female partners for the first sexual partnership (if that partnership isn't a cohabitation). These four models correspond to four categories of age at first sex: $12-19$, $20-29$, $30-37$, ≥ 38 years old. The predictors for these models were: birth year, lifetime number of partners at the time of interview, lifetime number of cohabitations at the time of interview, age at interview, age at first sex. We determined the age of the male partner of the first sexual partnerships using these models. The goal was to then use the age difference at first sex as a predictor for the age difference of subsequent non-cohabiting partnerships (thereby introducing a within-individual correlation).

Second, we fitted four regression models of mixture of two normal law, one logistic regression model, and one regression on the mean of a gamma distribution to estimate the age difference for non-cohabiting partnerships other than the first sexual partnerships. The logistic model estimates the odd that a woman has a male partner more than 15 years older or younger than her. If the age difference is greater than 15 years, the gamma regression estimates the age difference between the male and female partner. If the age difference is under 15 years, one of the three models of two normal law mixture estimates the age difference. These three models correspond to three categories of age at start of partnership: $12-24$, $25-34$, ≥ 35 years old. For all three models, predictors of the mean were the age at the start of partnership; predictors of the logistic regression model for the mixture were age difference between male and female partners at first sex, birth year, lifetime number of partners at the time of interview, age at interview, age at first sex.

B.4 Men's sexual history

B.4.1 Overview

Men's sexual history was dependent on the women's sexual history. Dates of cohabitation and of non-cohabiting partnerships in men were determined simultaneously with the sexual mixing with women. We proceeded in five steps: 1) Estimate an incomplete profile of sexual history for every man, 2) Identify who mixes with whom for cohabitation partnerships, 3) Complete the profile of cohabitation history of men, 4) Estimate an incomplete profile for non-cohabiting partnerships, 5) Identify who mixes with whom for non-cohabiting partnerships, 6) Complete the profile of sexual history for men.

B.5 Step 1: Incomplete profile of men's sexual history

Complete models formulations and illustrations of fits are given in Technical Appendix C.3.

B.5.1 Number of cohabitations

We estimated the number of cohabitations of each man given his birth year. We used data from NHSLS, NSFG surveys of 2002, 2006-2010, 2011-2013, and 2013-2015 on the number of cohabitations of men born between 1945 and 1999. Three logistic regression models were fitted to determine the probability of having had 0 cohabitation, 1 cohabitation, 2 cohabitations, and 3 or more cohabitations. A negative binomial regression model was fitted to determine the exact number of cohabitations for men with 3 or more cohabitations (from 3 to 8 cohabitations). In all four models, the only predictors were age at interview and birth year.

B.5.2 Approximate length of cohabitations and age at first cohabitation

The lengths of cohabitations were categorized as]0,1] year,]1,5] years,]5,10] years, > 10 years. For every cohabitation, we determined the approximate length of cohabitation as one of these four categories. To do this, every man of the model was matched for birth year and number of cohabitations with a male participant of the NSFG or NHSLS. Lengths of cohabitations and age at first cohabitation were derived directly from the matched male participant.

B.5.3 Number of non-cohabiting sexual partnerships

We estimated the number of non-cohabiting sexual partnerships given the incomplete cohabitation history of men. We proceeded as we did with women, first splitting men in 6 classes: 0 partner, $1-3$ partners, $4-10$ partners, $11-20$ partners, $20-40$ partners, ≥ 41, partners. We fitted five logistic regression models to determine class membership, four negative binomial regression models and one power law regression to determine the exact number of partners within each class. The power law was used to estimate the number of partners of men with more than 40 partners without using any predictors. For all other models, the predictors were birth year, number of cohabitations, lengths of cohabitations, and age at interview.

B.5.4 Race of female cohabitation partner

We used data on the race of each female cohabitation partner in NHSLS, NSFG 2002, 2006-2010, 2011-2013, 2013-2015 to fit three logistic regression models. Race is categorized as Non-Hispanic White, Non-Hispanic Black, Hispanic and Others. The first model is used to estimate the probability that the partner is White. The second model estimates the probability that the partner is Black conditional on the partner not being White. The last model estimates the probability that the partner is Hispanic conditional on the partner being neither White nor Black. For all three models, predictors were the birth year, duration of cohabitation, lifetime number of partners at the time of interview, lifetime number of cohabitations at the time of interview, number of cohabitations at the time of cohabitation initiation.

B.5.5 Age of female cohabitation partner

We used data on the age of the female cohabitation partners of men in NHSLS, NSFG 2002, 2006-2010, 2011-2013, 2013-2015 to fit four regression models of mixture of two normal law to estimate the age difference: age of man minus age of woman. These four models correspond to four categories of age at cohabitation: $12-19$, $20-24$, $25-34$, ≥ 35 years old. The predictors for these models were: age at interview, lifetime number of partners at the time of interview, number of previous cohabitations at the time of cohabitation, birth year, length of cohabitation. The estimation of the age of the female partner is only necessary to determine if the female partner is outside the model population (see below B.7).

B.6 Older birth cohorts: 1850-1944

Complete models formulations and illustrations of fits are given in Technical Appendix C.4.

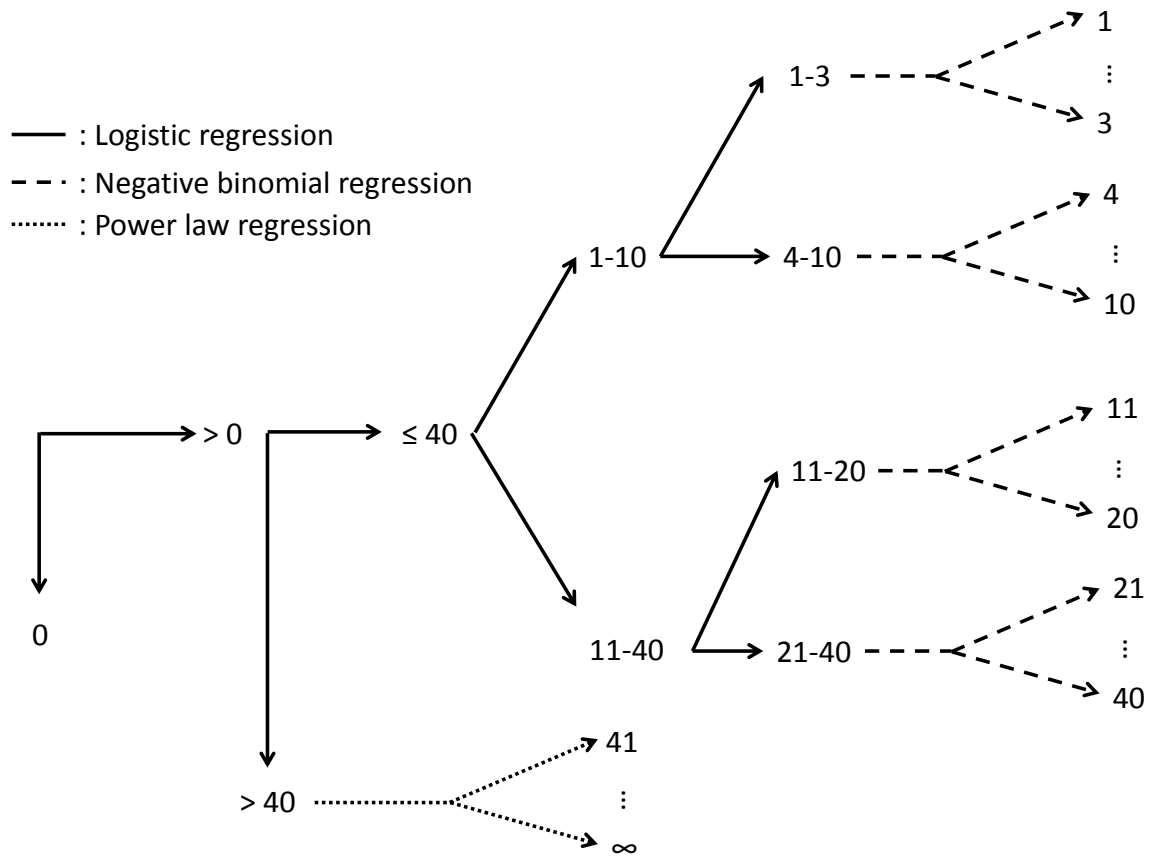


Figure B.5 – **Process of estimating the number of non-cohabiting sexual partnerships among men.** Men were split into six classes using logistic regression (0, $1-3$, $4-10$, $11-20$, $21-40$, ≥ 41). Negative binomial regressions were used to estimate the number of non-cohabiting sexual partners within each of the first three classes. A power law regression was used for estimation within the last class (≥ 41).

B.6.1 Overview

We are primarily interested in cohorts born between 1945 and 1999. However, individuals born close to 1945 will have sexual partners born before 1945. Hence, if we are interested in simulating the transmission of STIs for birth cohorts between 1945 and 1999, we still need to take in account the sexual activity and transmission of cohorts born before 1945. This poses the problem of "closing" the population. Typically, this is done by assuming that the population is initially stable. In our case, this would mean that the 1945 population is stable as far as sexual activity is concerned. However, there are strong evidences that changes in sexual behavior were occurring in the beginning of the 20th century, much earlier than the birth cohort of 1945. Therefore, we chose rather to assume a stable population before 1900 with changes in sexual behavior occurring from the birth cohort of 1900 up until the birth cohort of 1999. We thus extended the range of the birth cohorts we consider from 1945-1999 to 1850-1999, with the birth cohorts between 1850-1900 being stable (the same). This additional period from 1850 to 1900 is included to allow transmission to reach an equilibrium state before the

beginning of the 20th century.

To determine the sexual activity of the birth cohorts between 1900 and 1944, we used data on the lifetime number of sexual partners (since age 18) and the age at first marriage from the GSS between 1989-2016. Since the GSS has no age limit, unlike the NSFG and NHSLS, it is possible to have partial data on the sexual history of men and women born between 1900 and 1944. Sexual activity for the birth cohorts before 1900 was assumed to be identical to the birth cohort of 1900.

B.6.2 Lifetime number of partners of men and women born between 1850-1944

Using the data from GSS on the number of sexual partner since 18 years old, we fitted models to determine the lifetime number of partners like the models described in B.2.4, but we only use year of birth as predictor.

B.6.3 Age at first cohabitation of men and women born between 1850-1944

Using the data from GSS on age at first marriage, we fitted a multinomial model to determine the age at first cohabitation categorized as $[0, 20]$ years, $]20, 25]$ years, $]25, 30]$ years, > 30 years. The predictors of the model were year of birth, and lifetime number of partners since 18 years old. It is assumed in this step that age at first marriage coincides with age at first cohabitation for these birth cohorts.

B.6.4 Completion of the sexual history of men and women born between 1850-1944

We completed the information on the sexual history of men and women born before 1945 by matching them with a participant from NHSLS born between 1945 and 1955, with the same number of partners after 18 years old, the same gender, and the same age at first cohabitation. We then took the number of partners before 18 years old of the match in the NHSLS as the number of partners before 18 years old of the individual of the model to obtain his total lifetime number of partners. We applied a correction to account for non-intercourse partners in GSS (see Technical Appendix C.3.2). This step was omitted for women. For both men and women, we used the sexual history of the match in NHSLS to complete the missing information: characteristics of sexual partners, age at first sex, number and lengths of cohabitations.

B.7 Step 2: Who mixes with whom: cohabitations with men inside the model population

B.7.1 Overview

For every cohabitation recorded in the sexual history of women, we determined which man was the partner. For cohabitations occurring with a man inside the modelled population (i.e., white man between 12 and 44 years old and born between 1850-1999), the following characteristics were known and recorded in the sexual history of the woman: 1) length of cohabitation, 2) age of male partner 3) date at start of cohabitation. Using this information, we identified all men that were eligible to have had this cohabitation and sampled one among them.

Eligibility of male partners

For a given cohabitation of a woman, we identified men who were not in cohabitation at the time it occurred, who had an upcoming cohabitation in their sexual history, and who had the same age at the time of the cohabitation as the age of the male partner recorded in the woman's sexual history. Then, we considered the following recorded characteristics of the upcoming cohabitations of the potential male partners: 1) length of cohabitation, 2) age at cohabitation (if known). We selected only men for whom the characteristics of the upcoming cohabitation matched the characteristics of the given cohabitation of the woman. If no men could be found, we progressively increased the range of eligible ages for the male partner and then of eligible lengths of cohabitation until at least one man could be found.

B.7.2 Sampling of the male partner using lifetime number of partners

To sample among the eligible men, we split them into four classes of lifetime number of partners: $1-4$, $5-10$, $11-20$, ≥ 20. Then, we estimated the probabilities that the male partner belongs to each of the four classes. To do this, we used data on all cohabitations declared by men in NHSLS, NSFG 2002, 2006-2010, 2011-2013, 2013-2015 to fit three logistic regression models to determine class membership of the male partner of a given cohabitation. For all three models, the predictors were: the difference in age between the partners, the age of the male partner at the time of cohabitation initiation, the duration of cohabitation, and the year of birth. Once the class of the male partner was determined, we sampled a random man among all the remaining eligible men of that class.

B.7.3 Step 2: Who mixes with whom: cohabitations with individuals outside the model

For cohabitations of women with non-white men, men born before 1850, men born after 1999 or men older than 44 years, we simply determined the level of sexual activity of the male partners by using the logistic regression models described in section B.7.2. For instance, if a woman cohabits with a black man, we used the probabilities of class membership predicted for a white man partner with otherwise the same characteristics as the black man.

Furthermore, for every cohabitation of men after the first, we had to determine whether the next cohabitation was with a woman inside the model. To do this, we needed to impute first the age at the start of next cohabitation. To do this, we fitted models analogous to those fitted for females in B.2.3. Once the age at the start of the next hypothetical cohabitation was determined, we used the models described in B.5.4 and B.5.5 to determine the age and race of the female partner. If the hypothetical female partner was outside the model, we determined the level of sexual activity of the female partner as explained previously for male partners. However, we used directly the females in the model to fit the three logistic regression models to determine class membership ($1-4$, $5-10$, $11-20$, ≥ 21 lifetime partners) of the female partner of a given cohabitation of a man. For all three models, the predictors were: the difference in age between the partners, the age of the male partner at the time of cohabitation initiation, the duration of cohabitation, and the birth year.

B.8 Step 3: Complete the profile of men's cohabitation history

With the information provided in Step 2: Who mixes with whom, we completed the cohabitation history of all men. Importantly, it was possible for men to have a cohabitation(s) recorded in their

history with a female partner inside the population, but that no female partner was actually assigned to this cohabitation. The reason for this is that the number of cohabitations reported by men exceeds the number of cohabitations reported by women. Hence, some cohabitations of men had no female partner assigned to them. These cohabitations that remained unassigned after Step 2 were removed from the cohabitation history of the men and were instead converted to non-cohabiting sexual partnerships. As a result, the distribution of the number of cohabitations for men changed after Step 2: the proportion of men with 3 or more cohabitations declined.

B.9 Step 4: Estimate an incomplete profile of non-cohabiting partnerships

B.9.1 Estimating the age at first sex for men

We used data on the age at first sex for men in NHSLS, NSFG 2002, 2006-2010, 2011-2013, 2013-2015 to fit two logistic regression and three negative binomial regression models to determine the gap in years between the age at first sex and age at first cohabitation. The logistic regression models were used to classify the gap as either low ([0,3] years), average (]3,9] years) or high ($\geq$ 9 years). The negative binomial models were used to determine the exact gap in years for each of the three classes. For all models, predictors were the lifetime number of partners at interview, age at first cohabitation, birth year, age at interview.

B.9.2 Estimating the period of non-cohabiting sexual partnerships

We proceeded as described for women in B.2.4. Thus, we attributed each of the non-cohabiting partnership to one of the 17 possible periods delimited by the beginning and end of each cohabitation with sexual initiation being the starting point. Men were split in four classes according to their number of non-cohabiting partners ($1-3$, $4-8$, $9-18$, ≥ 19) and we fitted a model for each class using the NHSLS data:

$$P[t_i = k] = \frac{W_k}{\sum_j W_j}$$

where t_i is the period in which occurred the i^{th} non-cohabiting partner, and W_j is the weight of period j. Each weight W_j is given by a function:

$$log(\frac{W_j}{N_j}) = f(\beta_j, Predictors)$$

where N_j is the number of months in period j and $f(\beta_j, Predictors)$ is a piecewise-linear function with parameters to be estimated. This model was used to attribute each non-cohabiting partnership to a specific period.

B.9.3 Estimating the months of non-cohabiting sexual partnerships occurring in a period outside a cohabitation partnership

After having classified sexual partnerships in one of the 17 periods described in B.2.4, we estimated the month during which every partnership occurred. For partnerships occurring in a period outside a cohabitation partnership (i.e., odd period), we sampled the months of occurrence of every partnership through a random sampling with replacement.

B.9.4 Estimating the months of non-cohabiting sexual partnerships occurring during a cohabitation partnership

We used the exact same procedure as in B.2.4 for women.

B.9.5 Men who pay for sex

To determine who has paid for sex, we used data from the GSS in which men are asked if they have ever paid for sex in their lifetime. We fitted a logistic regression model to these data using as predictors the lifetime number of partners at interview and age at interview. We thus had a subpopulation of men who pay for sex (MWPS).

B.9.6 Race of non-cohabiting female partners

As in B.2.4, we used data on the race of lifetime non-cohabiting partners only available in the NHSLS and demographic data on the race composition of the US population by birth cohort from decennial census to fit a model of preferential mixing. The model is used for extrapolation because the NHSLS data do not cover the younger birth cohorts. The formulation of the model is:

$$P_{i,y}[Race = k] = \frac{N_{k,y}W_{i,k}}{\sum_j N_{j,y}W_{i,j}}$$

with j in {White, Black, Hispanic, Others}, y is the calendar year the partnership occurred, i denotes whether an individual has paid for sex, $N_{k,y}$ is the population size of men of race k at year y, $W_{i,k}$ is a preferential weight of white men towards women of race k, and $P_{i,y}[Race = k]$ is the probability that a new female partner is of race k for a partnership occurring year y. The weights $W_{i,k}$ are modelled as:

$$W_{i,k} = exp(\alpha_{i,k} + \beta_{i,k} log(LP))$$

where LP is the lifetime number of partners and $\alpha_{i,k}$,$\beta_{i,k}$ are parameters to be fitted. Hence, there are two different sets of preferential weights depending on whether a man has paid for sex. This replaces the intra-individual correlation parameter that was included in the model for women in B.2.4.

B.9.7 Age of non-cohabiting female partners

We used data on the age of the female non-cohabiting partners of men in NHSLS, NSFG 2002, 2006-2010, 2011-2013, 2013-2015 to fit three regression models of mixture of two normal laws to estimate the age difference: age of man minus age of woman. These three models correspond to three categories of age at interview: $12-17$, $18-29$, ≥ 30 years old. The predictors of these models were: age at interview, lifetime number of partners at interview, lifetime number of cohabitations at interview, age at start of partnership.

B.9.8 Balancing non-cohabiting partnerships

In the national US surveys, the total number of non-cohabiting partnerships reported by men far exceeds the number of non-cohabiting partnerships reported by women. This should not be possible in a closed population, and may be a consequence of 1) reporting bias (e.g., due to social desirability), 2) under-sampling and under-reporting of core-groups of highly sexually active women (e.g., sex workers), 3) the population is not closed (e.g., men having sex outside of the US).

Scenario 1: Under-sampling and under-reporting of the core group
In Scenario 1, we assumed that the reason for the discrepancy in the reported number of sexual partners is the under-sampling of US female sex workers or other women with extremely high number of partners, the absence of foreign female sex workers in the surveys populations, and the under-reporting of sex workers and high-risk women who do get sampled in the survey. In fact, it is unlikely sex workers can accurately report their lifetime number of partners if it exceeds thousands of partners, and that value would be right-censored in most databases (the maximum possible value is 50 for NSFG after 1995). Hence, in this scenario, partnerships in excess reported by men are assumed to be formed with small groups of highly sexually active women either inside or outside the US.

In the model population, we created for the purpose of balancing partnerships a population of highly sexually active women, which includes sex workers either inside or outside the US population. To do this, we sampled among the existing women in the model a subpopulation of 40 women with at least 50 non-cohabiting partnerships. We duplicated these 40 women for every year of birth, generating groups for every year of birth from 1850 to 1999. Next, we compared the non-cohabiting partnerships of men with the non-cohabiting partnerships of women and calculated the excess. We first classified the non-cohabiting partnerships according to 1) the year of birth of the male partner (ten-years categories), 2) the age of the male partner ($12-15$, $16-19$, $20-24$, $25-34$, $35-44$ years old), and 3) the age difference between the male and female partners (≤ -3, $]-3,1]$, $]1,3]$, >3 years). Then, within each class thus defined, we calculated the difference between the number of partnerships of men and the number of partnerships of women. The result is the total number of partnerships in excess. As per our assumption, the partnerships of men in excess have to be formed with a member of the high-risk group (High-risk women: HRW). We then determined which men formed partnerships with the HRW (men-with-high-risk-women: MWHRW). We automatically included MWPS among MWHRW, as determined in B.9.5. We thus first sampled among partnerships of MWPS until we reached the number of partnerships in excess. Since this was not possible due to an insufficient number of MWPS, we increased the number of MWHRW by sampling among men who do not pay for sex. We did this by taking in account the likelihood of each man to have had paid for sex as calculated with the model used in B.9.5. We also forced 10% of the men with the highest number of partnerships to be MWHRW. After this process, non-cohabiting partnerships of men have been classified according to whether they are formed with a member of the HRW. The number of partnerships of men that were not formed with a member of the HRW equals the number of partnerships of women.

Scenario 2: Men overestimate and women underestimate
There is evidence that men overestimate their number of sexual partners. This could be due to: 1) a higher frequency of "rounding up" the lifetime number of partner when estimating a high number (e.g., rounding up 26 partners to 30), 2) social desirability for men to have many partners (e.g., prestige associated with having many sexual partners), 3) inclusion of non-intercourse partners in the count (e.g., oral sex only partners). Similarly, there is evidence that women underestimate their number of sexual partners and the most plausible explanation for this would be social desirability. For both men and women, the reporting bias may have changed with the evolution of social norms and attitudes toward sex.

In Scenario 2, we assumed that men overestimating and women underestimating are together responsible in equal parts for about 50% of the discrepancy in lifetime number of partners. The remaining of the discrepancy is explained by the under-sampling and under-reporting of a core group of women as explained in Scenario 1.

To determine which men overestimated their number of lifetime partners, we used data from NSFG on sex partners that include oral and anal sex partners asked in the Audio Computer-Assisted Self-Interviewing (ACASI) section to fit a model identifying men who report a lower lifetime number of partners in the ACASI section compared to the Computer-Assisted Personal Interview (CAPI) section. A lower number on the ACASI section, which includes more sexual acts, is assumed to be caused by either a more precise (re)estimation of the lifetime number partners or a more truthful answer under less social pressure. We also fitted a negative binomial model of the difference between the ACASI report and the CAPI report to estimate the magnitude of the overestimation. Using these two models, we identified men from the model population with overestimated lifetime number of partners and we determined their real lifetime number of partners. The extra partnerships were removed randomly (only non-cohabiting partnerships were removed). Since we could not thus remove enough partnerships, we reselected new men by sampling according to their likelihood of having overestimated their lifetime number of partners iteratively until we reached the quota (25% of the discrepancy).

We proceeded analogously for women, using this time a positive difference between the ACASI report of the lifetime number of intercourse partners in NSFG 95 and the CAPI report. The resulting correction of the distribution of lifetime number of sexual partners is presented in Technical Appendix C.8.

Scenario 3: Men overestimate
In scenario 3, we assumed only men were overestimating. We proceeded otherwise as in Scenario 2. Around a third of the discrepancy in lifetime number of partners was explained by men overestimating. The remaining excess of partnerships was absorbed by the HRW.

B.10 Step 5: Who mixes with whom: non-cohabiting partnerships

B.10.1 Overview

The dates of non-cohabiting partnerships were recorded independently in both men's and women's sexual history. To determine sexual mixing, these records of men and women need to be merged. For partnerships in which both partners are inside the population, we determined the exact date of the sexual partnerships from the women's records, and we allowed the dates of the partnerships recorded in the men's sexual history to be flexible within a period (defined using the start and end of each cohabitation). There is an exception to this for first-sex partnerships of men.

We distinguished two types of non-cohabiting partnerships recorded in the sexual history of women: 1) partnership with a man inside the model population: a white man between 12 and 44 years old and born between 1850-1999, 2) partnership with a man outside of the model population.

For every non-cohabiting partnership recorded in the sexual history of women with a man inside the model population, we determined which of the men the partner was. For every such partnership, the following characteristics were known and recorded in the woman sexual history: 1) age of male partner 2) date of partnership. Using this information and the information recorded in men's sexual history, we first determined the level of sexual activity of the male partner. We then identified all men that were eligible to have had this non-cohabiting partnership and sampled one man among them.

For partnership with men outside the model population, we only identified the level of sexual activity of the male partner.

B.10.2 Level of sexual activity of the male partner

We classified the men into four classes of lifetime number of partners: $1-4$, $5-10$, $11-20$, ≥ 21. Then, we estimated the probabilities that the male partner belongs to each of the four classes. To do this, we used information on every non-cohabiting partnership recorded in all the men's sexual histories. We fitted three logistic regression models to determine class membership of the male partner of a given non-cohabiting partnership. For all three models, the predictors were: the difference in age between the partners, the age of the male partner at date of partnership, and the birth year.

B.10.3 Eligibility of male partners

For a given non-cohabiting partnership of a woman, we first identified, for every man, during which of the 17 periods (see B.2.4) the partnership occurred. Then, we identified all men who had a non-cohabiting partnership with a woman inside the model (outside HRW) recorded during the identified period, and who had the same age at the time of the partnership as the age of the male partner recorded in the woman's sexual history. We sampled randomly a man among all these men. For the sampling, we used as weights for each man the number of non-cohabiting partnership still unassigned to a woman within the period divided by the number of months remaining in the period. If no men could be found, we randomly picked another class of sexual activity for the man. If still no men could be found, we progressively increased the range of eligible ages of the male partner until at least one man could be found.

B.10.4 Non-cohabiting partnership recorded in the sexual history of men

We distinguished four types of non-cohabiting partnerships for men: 1) non-paid partnerships to be assigned through the sexual history of women as described in B.10.3, 2) partnerships with a woman outside the model but not in the HRW group, 3) partnerships to be assigned with the HRW group, 4) first-sex-partnerships that occurs with women inside the model population (outside HRW).

For first-sex-partnerships, the exact date of the partnership is informed by empirical data (i.e., the age at first sex), so we aimed to preserve the exact date of the first partnership for these men. This is not the case for partnerships in 1) that are assigned through the sexual history of women, as the date of these partnerships may end up not exactly fitting the recorded date in the men's sexual history. Hence, for first-sex-partnership that occurs with women inside the model population, we identified eligible women and picked one. For 2) partnerships with HRW, we identified eligible women among HRW and picked one. For 3) non-paid partnerships with women outside the model population, we only determined the level of sexual activity of the female partner.

Eligibility of female partners for first-sex-partnerships

For a given non-cohabiting first-sex-partnership of a man, we identified all woman who had a non-cohabiting partnership recorded at the exact date of the partnership with a man inside the model, and who had the same age at the time of the partnership as the age of the female partner recorded in the man's sexual history. Among the remaining eligible women, we chose the one for whom the recorded age of the male partner was the closest to the actual age of the male. If no woman could be found, we changed the first-sex-partnership to be with a HRW.

Eligibility of HRW partners

For a given non-cohabiting partnership of a man with a HRW, we identified all HRW who had a non-cohabiting partnership recorded at the exact date of the partnership with a man inside the model population, and who had the same age at the time of the partnership as the age of the female partner recorded in the man's sexual history. Among the remaining eligible women, we chose one randomly. If no woman could be found, we progressively increased the range of eligible ages of the female partner until at least one woman could be found.

Level of sexual activity of the female partner outside the model population

For partnerships with female partners outside the model population but not in the HRW group, we estimated the probabilities that the female partner belonged to each of the four classes of sexual activity. To do this, we used information on every non-cohabiting partnership recorded in the women's sexual history. We fitted three logistic regression models to determine class membership of the female partner of a given non-cohabiting partnership. For all three models, the predictors were: the difference in age between the partners, the age of the male partner at the date of partnership, and the birth year. Thus, if a man had a partnership with a non-white female, it was assumed that the female partner had the same level of sexual activity as if she had been white. If a man had a partnership with a women aged more than 44 years old, it was assumed that the female partner had the same level of sexual activity as a 44 years old female would have had (a 44 years old female with a younger man, because the age difference between partners was kept the same).

B.11 Step 6: Complete the profile of sexual history for men

For partnerships with females inside the model population except the first one, the dates of partnerships are given by the female partner's sexual history. For all other partnerships, the dates remain the same as the ones determined in **Step 4**.

B.12 Individual-based dynamic model of STI transmission

B.12.1 Overview

The model follows a Susceptible $\rightarrow$ Infected $\rightarrow$ Recovered/Susceptible structure. Hence, individuals are born susceptible and can acquire the infection through a sexual partnership with someone infected. Once infected, the duration of infection is determined by an exponential random variable with mean equals to the average duration of infection. Upon clearing the infection, the individual may acquire natural immunity according to a probability of acquiring natural immunity. Else, the individual becomes susceptible again after clearing the infection. There are two sites of infection: the genital and oral sites.

B.12.2 Transmission and partnerships

All partnerships are assumed to be instantaneous for the purpose of transmission. This means that once a partnership occurs between an infected and a susceptible, a Bernoulli trial is made to determine whether transmission occurs according to a probability of transmission per instantaneous partnership. Transmission to the genital site occurs in a partnership in which the infected individual has a genital infection and the susceptible individual has no genital infection. Transmission to the oral site occurs in a partnership in which the infected individual has a genital infection and the susceptible individual has

no oral infection. We are thus modeling two transmission pathways: genital → genital and genital→ oral.

As before, we distinguished two types of partnerships: 1) partnerships between two individuals within the model population (white, aged 12-44 years old, born between 1850 and 1999) 2) partnerships in which at least one of the individual is outside the model population.

For the partnerships of type 2), we cannot directly tell whether the partner outside the model population is infected or not. However, we have recorded the following information on every partner outside the model population: age, year of birth, and class of sexual activity ($1-4$, $5-10$, $11-20$, ≥ 21 lifetime number of partners). Based on these characteristics, we estimated the probability that the partner outside the model population was infected. To do this, we used the prevalence of infection in the subgroup to which belongs the partner. For a partner aged more than 44 years old, we used the prevalence of infection in the same birth cohort and sexual activity class as the partner but at age 44 years old. We are thus assuming that prevalence of infection remains stable after 44 years of age. For a partner aged less than 12 years old, we used the prevalence among those aged 12 years old. For a partner that is not white, we used the prevalence of infection among whites in the same birth cohort, same age, and same sexual activity class as the partner. Hence, we are assuming that given a birth cohort, age and level of sexual activity, the prevalence of infection is the same regardless of the race. In all previous cases, the prevalence is calculated only among individuals who will form a partnership during the month of occurrence of the type 2) partnership to avoid selection bias.

B.12.3 Scenarios of transmission

We considered three scenarios regarding transmission of HPV. In the first scenario *uni-site*, HPV is only transmitted to the genital site. Hence, there is a single parameter of transmission: the probability of genital→genital transmission per partnership. In the second scenario *multi-site with oral sex*, HPV can be transmitted to both the genital and oral sites. There are two parameters of transmission, the probability of genital→genital transmission per partnership, and the probability of genital→oral transmission per partnership in which oral sex is practiced. Hence, in this scenario, the only way to acquire oral HPV is to perform oral sex on someone infected at the genital site. In the third and last scenario *multi-site without oral sex*, HPV can be also be transmitted to the genital and oral sites. There are again two parameters of transmission, the probability of genital→genital transmission per partnership, and the probability of genital→oral transmission per partnership (regardless of whether oral sex is practiced in the partnership). Hence, in this scenario, oral HPV is acquired through sexual intercourse with someone infected at the genital site.

B.12.4 Calibration of transmission probabilities

Calibration was performed separately for each of the possible combination of scenarios (*Uni-site, multi-site with oral sex, multi-site without oral sex, High-risk group, Report bias men & women, Report bias men only*). Calibration was also performed separately for the probability of genital→ genital transmission and the probability of genital→oral transmission. This was possible because the former probability generates the genital HPV16 prevalence while the latter probability generates the oral HPV16 prevalence. Calibration was performed by the bisection algorithm, testing values

between 0 and 1 for the probabilities of transmission and minimizing the mean absolute error over 5 simulations.

B.13 Transition to oropharyngeal and cervical cancers

We used SEER incidence data to fit the parameters of transition from infection to cervical and oropharyngeal cancers for each of the 20 simulations per scenario (180 total simulations). We first obtained the mean and standard deviation of the time from infection to cervical cancer. To do this, we used as targets the incidence of HPV16-positive cervical SCC between 2004 and 2008 in the following age groups: [20,25[, [25,30[, [30,35[, [35,40[, [40,45[, [45,50[years. We performed the calibration for each of the 20 simulations by testing parameter values of the mean time from infection to cancer between 20 and 50 years, and values of the standard deviation of the time from infection to cancer between 1 and 20. We then minimized the mean squared error. We then fitted the probability of a newly acquired infection transitioning to cervical cancer to the incidence of HPV16-positive cervical SCC between 1973 and 1975, during which the impact of cervical cancer screening would be the least (1973 is the minimum year of the SEER data). To obtain the targets of HPV16-positive cervical incidence, we assumed a proportion of HPV16-positive cervical SCC of 0.54 based on a meta-analysis[223]. The proportion of HPV16-positive cervical SCC was assumed to be constant, i.e., independent from calendar time.

To obtain the mean and standard deviation of the time from infection to oropharyngeal cancer, we used as targets the incidence of HPV16-positive oropharyngeal SCC between 2004 and 2008 in the following age groups: [20,25[, [25,30[, [30,35[, [35,40[, [40,45[, [45,50[, [50,55[, [55,60], [60,65[. We performed the calibration as described for cervical cancer. We then fitted the probability of an infection transitioning to oropharyngeal cancer to the incidence of HPV16-positive oropharyngeal cancer between 1984 and 1985. The proportion of HPV16-positive SCC between 2004 and 2008 was obtained from a meta-analysis[3]. The proportion of HPV16-positive oropharyngeal cancer between 1984 and 2004 was obtained from a study[16] in which the proportion of HPV16-positive oropharyngeal SCC is estimated by DNA typing frozen cancer specimens between 1984 and 2004. We also extrapolated the results from this study by assuming that the proportion of HPV16-positive oropharyngeal SCC between 1973 and 1983 was the same as in 1984, and the proportion after 2004 was the same as in 2004. We thus have estimated the incidence of HPV16-positive oropharyngeal SCC from 1973 to 2015.

B.14 Projection of HPV16-positive oropharyngeal cancer incidence up to 2045

We estimated the incidence of HPV16-positive oropharyngeal cancer from 1985 to 2045. To obtain the incidence between 2015 and 2045, we restricted the population to men aged more than 50 years old. This is because younger men between 2015 and 2045 may not be in the model population which includes cohorts born before 2000. Because HPV transmission is simulated up to 2015, it was also necessary to extrapolate incidence of HPV infection between 2015 and 2045 in men. To extrapolate, we assumed that the age-specific incidence of HPV infection remained the same from 2015 to 2045 as in 2015. Of note, the great majority of oropharyngeal cancers in men aged more than 50 years old will stem from infection that were contracted before 2015 (and thus were not extrapolated) since the mean time from infection to cancer is around 40 years.

Appendix C

Technical appendix. Impact of changes in sexual behavior on past and future trends of HPV infections and related cancers: Models formulations and illustrations of fit

C.1 Women's sexual history for birth cohorts born between 1945 and 1999

	Definitions
Variables	
A	current age
T_0	time since first intercourse
T_1	time since start of current cohabitation
T_2	time since end of last cohabitation
A_0	age at first intercourse
A_1	age at start of current/last cohabitation
A_2	age at end of last cohabitation
D	duration of previous cohabitation
C	lifetime number of cohabitations
B	year of birth
Transformations	
t_A	piece-wise linear decomposition with knots at 14, 16, 18, 20, and 24 years old
r_A	piece-wise linear decomposition with knots at 16, 18, 20, 24, 30, and 37 years old
g_T	piece-wise linear decomposition with knots at $1/12$, 1, 3, 5, and 8 years
g_B	piece-wise linear decomposition with knots at 1950, 1963, 1975, and 1988
g_A	piece-wise linear decomposition with knots at 16, and 20 years old
g_D	piece-wise linear decomposition with knots at at 5 years
h_T	piece-wise linear decomposition with knots at 1, 3, 5, 8, and 15 years
h_A	piece-wise linear decomposition with knots at 16, 20, 25, and 30 years old
s_A	piece-wise linear decomposition with knots at 30 years old

Table C.1 – **Definitions of variables and transformations used to describe the models of the date of first intercourse and cohabitations**

C.1.1 Date of first intercourse

The model formulation for date of first intercourse is:

$$\log\left(\frac{1}{\lambda(A,B)}\right) = f(\beta_0,\beta,A,B)$$

where $\frac{1}{\lambda(A,B)}$ is the *age*- and *birth year*- dependent average time to first intercourse and f a transformation of *age* and *birth year*, and β is the vector of parameters.

$$f(\beta_0,\beta,A,B) = \beta_0 + \beta \cdot (t_A(A) \times g_B(B))$$

Figure C.1 presents the fit of the model to the NSFG data by birth year. Table C.1 gives the definition for the variables and transformations used for the models of the date of first intercourse and the dates of cohabitations.

C.1.2 Dates of cohabitations

The models formulations for dates of cohabitation are:

$$log\left(\frac{1}{\lambda(predictors)}\right) = f(\beta,predictors)$$

where λ is the rate of initiation or separation which depends on the predictors, β is the vector of parameters. Rates are applied to at-risk person-time. For instance, rates of initiation of second cohabitation are applied only to individuals who have not yet entered their second cohabitation but have left their first cohabitation.

There is an exception for the initiation of the first cohabitation to allow the cohabitation to occur simultaneously with first intercourse. Thus, we also fit a logistic model for the probability of initiating cohabitation when first intercourse occur.

$\boldsymbol{ModelAFS}$ **:**

$$logit\,(P) = \beta_0 + \beta \cdot (g_B(B) \times t_A(A_0))$$

First cohabitation: initiation

$\boldsymbol{ModelCOHINIT_1}$:

$$f(\beta_0,\beta,\boldsymbol{B},\boldsymbol{A}) = \beta_0 + \beta \cdot (g_B(B) \times r_A(A))$$

This model yields the number of cohabitations for each month of age. To determine the count of new cohabitations conditional on age at first sex, we used cox regressions.

$$log\,(\lambda(t,A,A_0)) = \lambda_0(t) + \beta \cdot (r_A(A) \times t_A(A_0))$$

We fitted one such model for four categories of birth year ($1945-1958$, $1958-1971$, $1971-1983$, $1983-1999$). This process was exceptionally done for the first cohabitation to "break down" the models in two parts because of memory issues when fitting directly the full model.

First cohabitation: separation

$\boldsymbol{Model\ COHSEP_1}$:

$$f(\beta_0,\beta,\boldsymbol{A_1},\boldsymbol{B},\boldsymbol{T_1}) = \beta_0 + \beta \cdot (h_T(T_1) \times g_B(B) \times h_A(A_1))$$

Separate models are fitted for three categories of age at first sex ($12-16$ years old, $17-18$ years old, and $19-44$ years old).

Second cohabitation: initiation

$\boldsymbol{Model\ COHINIT_2}$:

$$f(\beta_0,\beta,\boldsymbol{T_2},\boldsymbol{B},\boldsymbol{A_1},\boldsymbol{A_2}) = \beta_0 + \beta \cdot (g_T(T_0) \times B \times t_A(A_1) \times s_A(A_2))$$

Separate models are fitted for three categories of age at first sex ($12-16$ years old, $17-18$ years old, and $19-44$ years old) and three categories of birth year ($1945-1958$, $1958-1975$, $1975-1999$).

Second cohabitation: separation

$\boldsymbol{Model\ COHSEP_2}$:

$$f(\beta_0,\beta,A_1,B,T_1,D) = \beta_0 + \beta \cdot (h_T(T_1) \times h_A(A_1) \times B \times g_D(D))$$

Separate models are fitted for three categories of age at first sex ($12-16$ years old, $17-18$ years old, and $19-44$ years old) and three categories of birth year ($1945-1958$, $1958-1975$, $1975-1999$).

Third to eighth cohabitation: initiation

$\boldsymbol{Model\ COHINIT_3}$:

$$f(\beta_0,\beta,C,T_2,B,A_1,A_2) = \beta_0 + \beta \cdot (g_T(T_0) \times B \times t_A(A_1) \times s_A(A_2))$$

Separate models are fitted for three categories of age at first sex ($12-16$ years old, $17-18$ years old, and $19-44$ years old) and three categories of birth year ($1945-1958$, $1958-1975$, $1975-1999$).

Third to eighth cohabitation: separation

$\boldsymbol{Model\ COHSEP_3}$:

$$f(\beta_0,\beta,A_1,B,T_1,D) = \beta_0 + \beta \cdot (h_T(T_1) \times h_A(A_1) \times B \times g_D(D))$$

Separate models are fitted for three categories of age at first sex ($12-16$ years old, $17-18$ years old, $19-44$ years old) and three categories of birth year ($1945-1958$, $1958-1975$, $1975-1999$).

Model fit figures

Figure C.1-C.8 presents the fit of the model to the NSFG data by birth year.

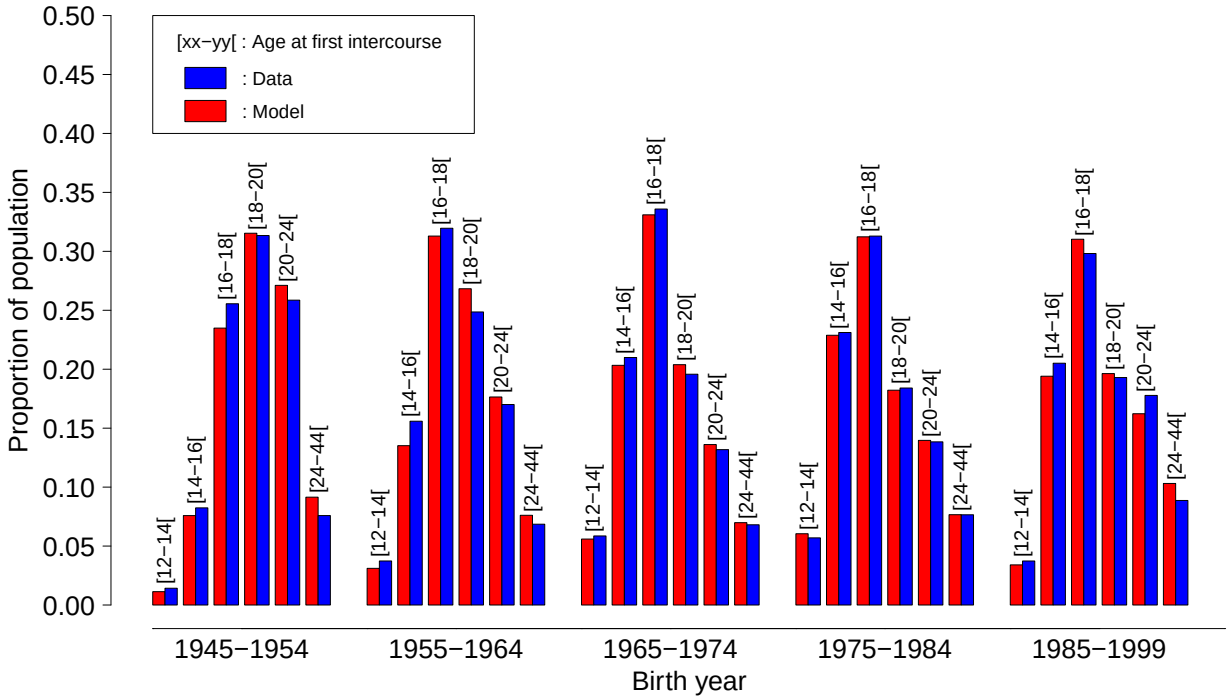


Figure C.1 – **Distributions of age at first intercourse for women by birth cohort: model simulations vs NSFG data**

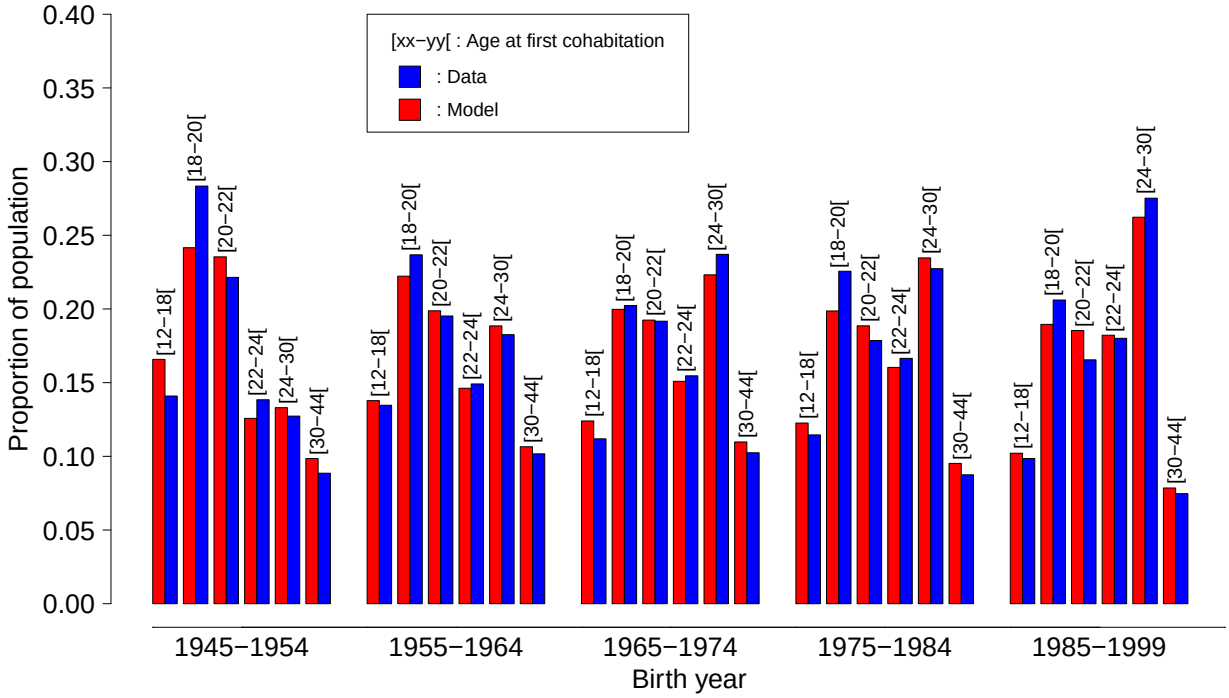


Figure C.2 – **Distributions of age at first cohabitation for women by birth cohort: model simulations vs NSFG data**

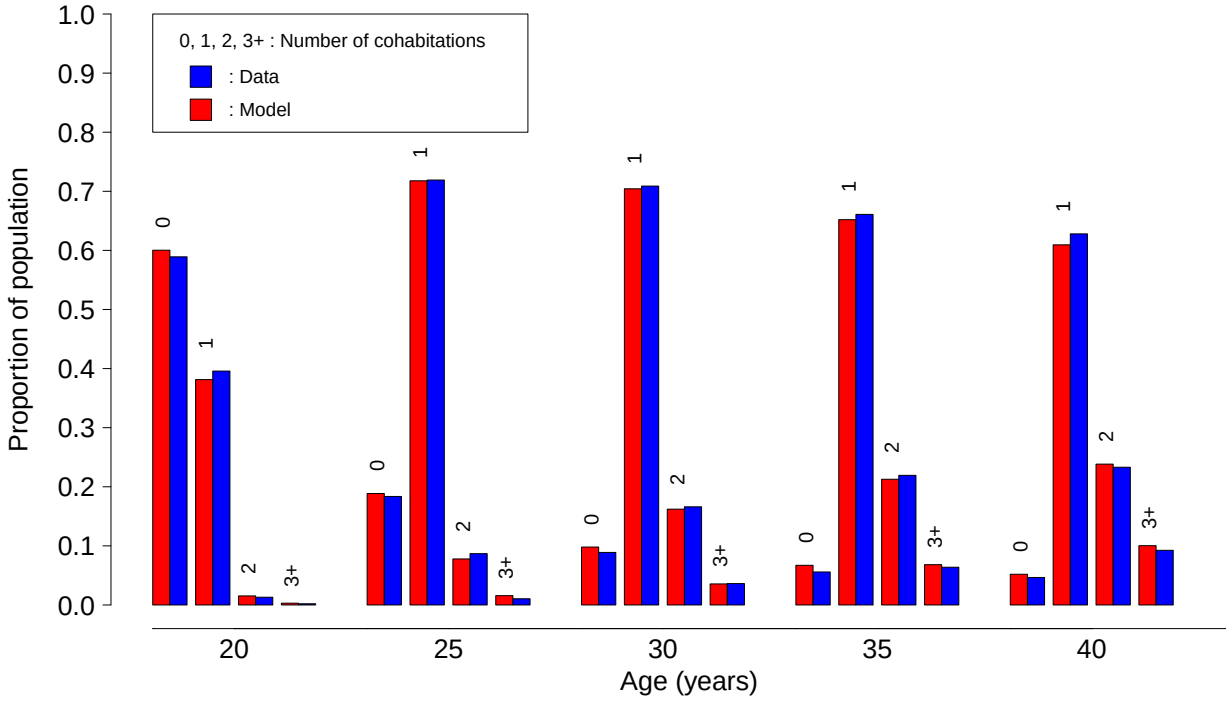


Figure C.3 – **Distributions of number of cohabitations by age for women born between 1945 and 1954: model simulations vs NSFG data**. 3+ means three or more.

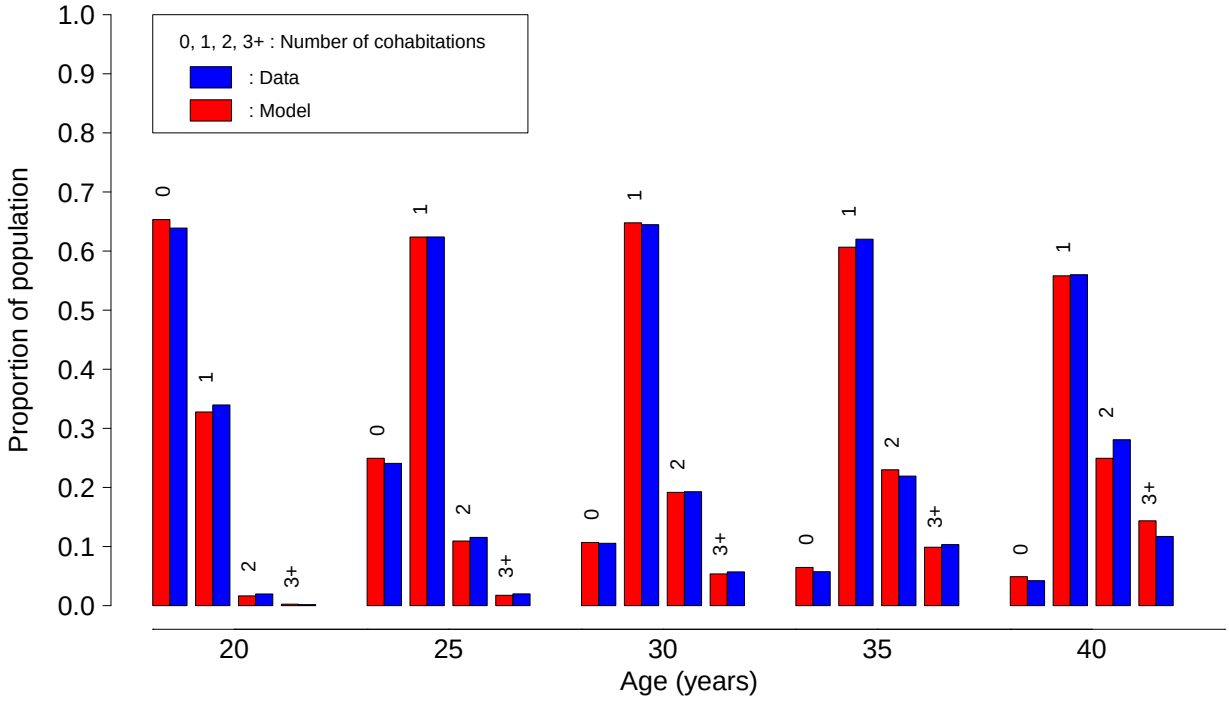


Figure C.4 – **Distributions of number of cohabitations by age for women born between 1955 and 1964: model simulations vs NSFG data**. 3+ means three or more.

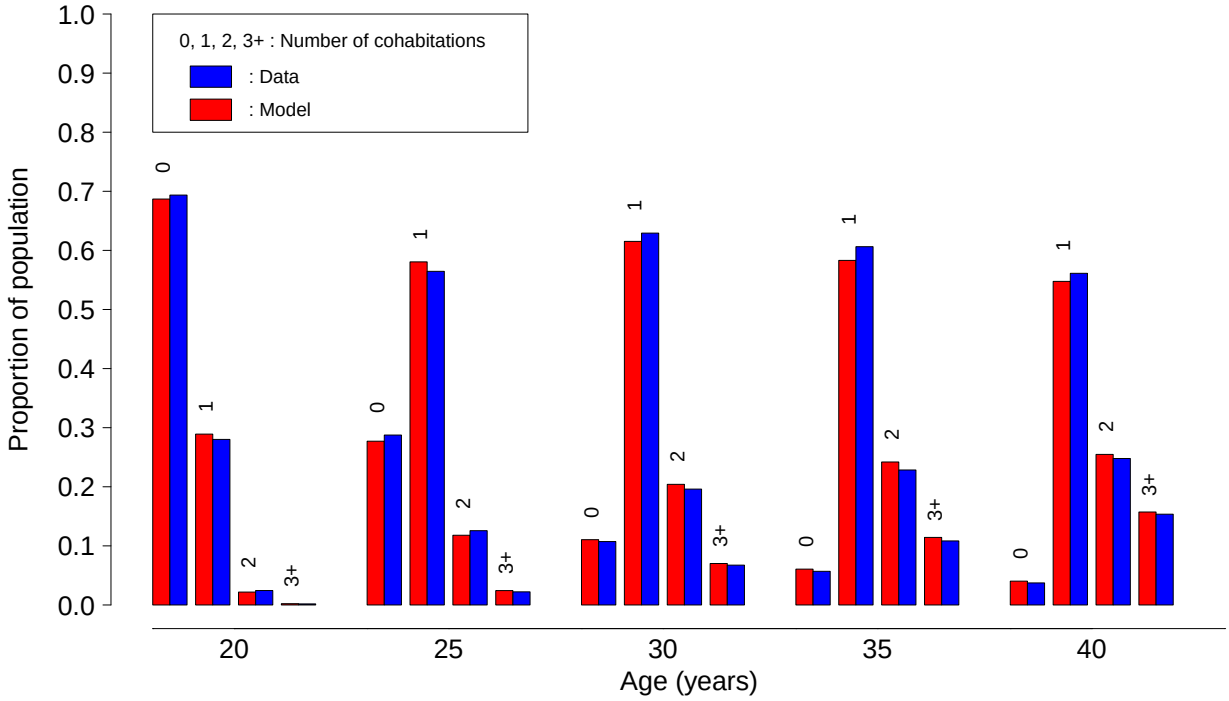


Figure C.5 – **Distributions of number of cohabitations by age for women born between 1965 and 1974: model simulations vs NSFG data**. 3+ means three or more.

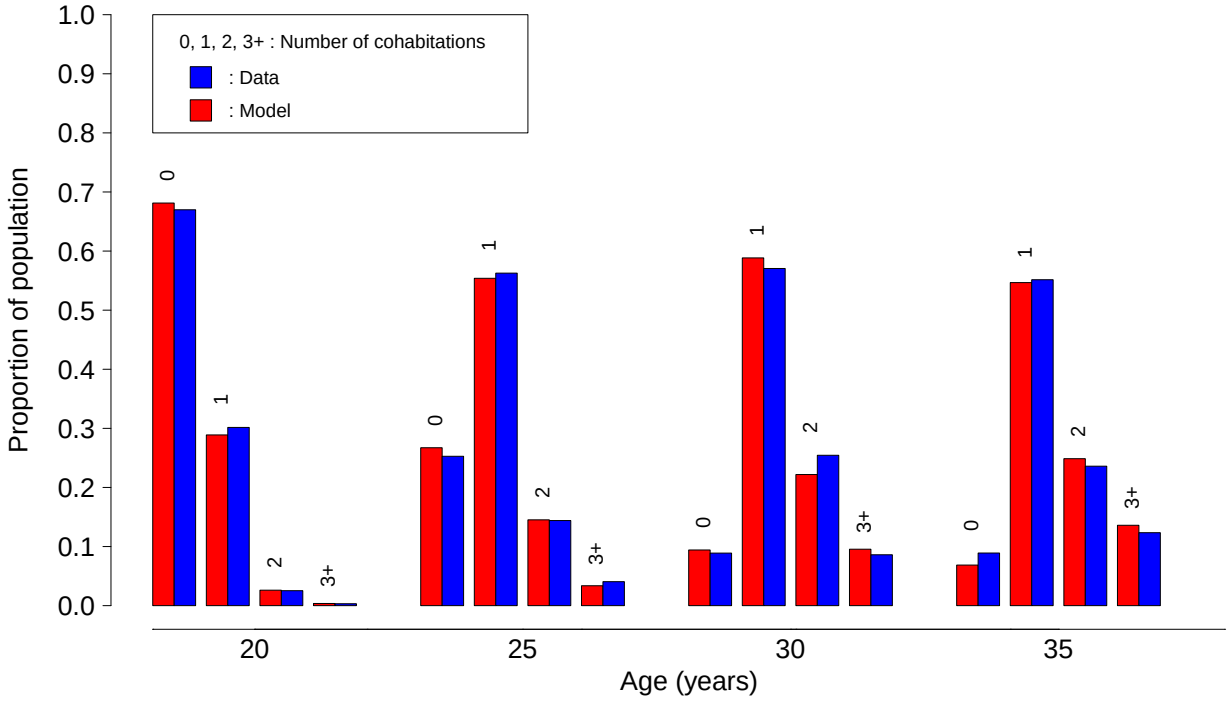


Figure C.6 – **Distributions of number of cohabitations by age for women born between 1975 and 1984: model simulations vs NSFG data**. 3+ means three or more.

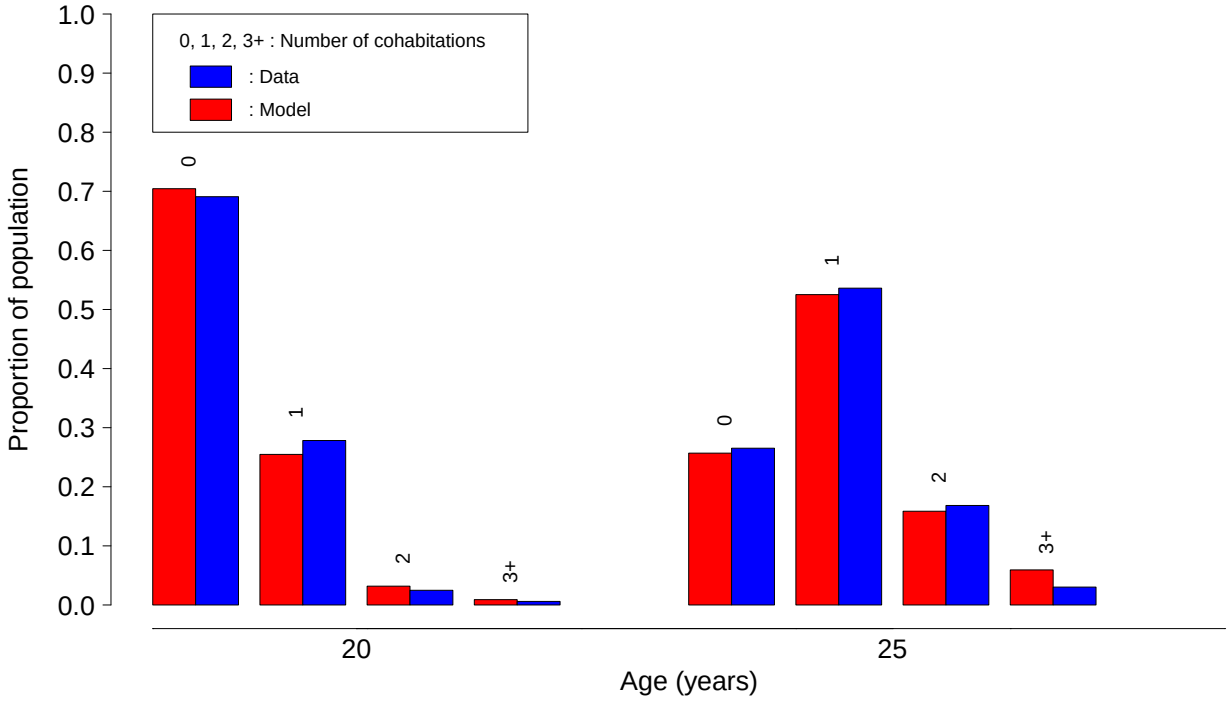


Figure C.7 – **Distributions of number of cohabitations by age for women born between 1985 and 1999: model simulations vs NSFG data**. 3+ means three cohabitations or more.

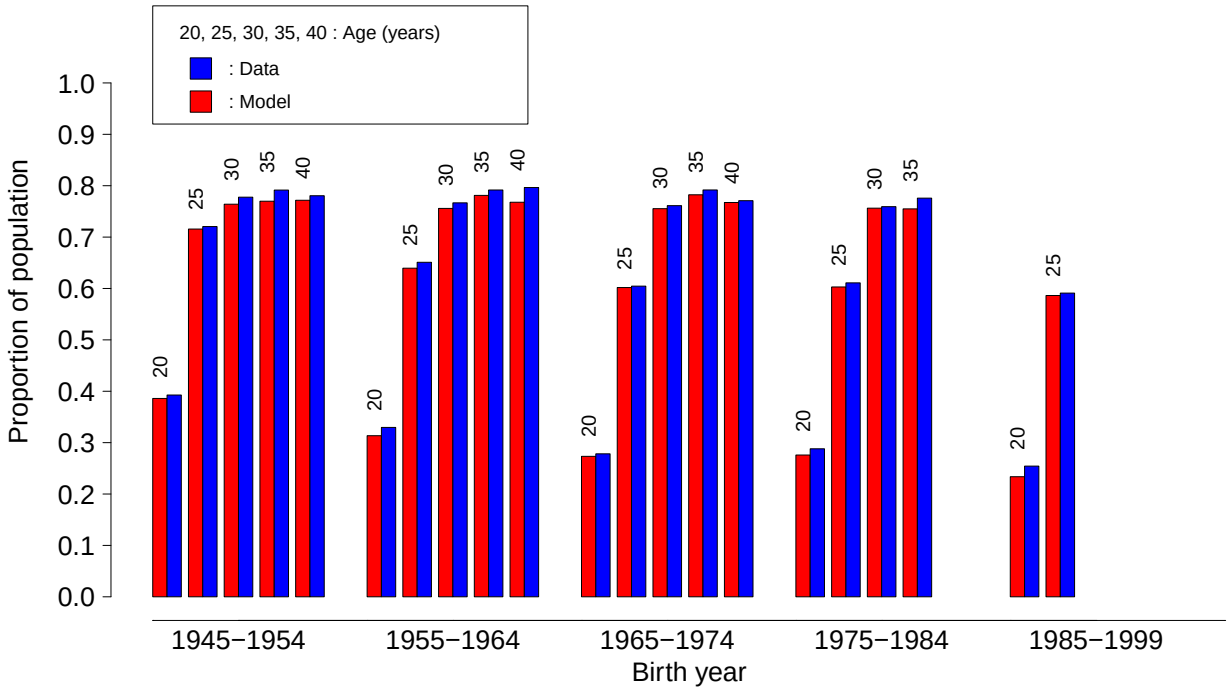


Figure C.8 – **Proportion of the population cohabiting for women by age and birth cohort: model simulations vs NSFG data**

C.1.3 Dates of non-cohabiting sexual partnerships

Lifetime number of non-cohabiting sexual partnerships

Table C.2 gives the definitions for the variables and transformations for the model of lifetime number of non-cohabiting sexual partnerships.

	Definitions
Variables	
A	current age
A_0	age at first intercourse
M_0	number of months between first intercourse and first cohabitation
M_1	number of months spent outside a cohabitation after the first cohabitation
	number of months spent outside a cohabitation after the first intercourse
M_2	weighted proportionally by age
	(a month at 20 years old is worth twice a month at 40 years old)
C	lifetime number of cohabitations
B	year of birth
Transformations	
t_A	piece-wise linear decomposition with knots at 20, 25, 30, and 38 years old
r_A	piece-wise linear decomposition with knots at 14, 16, 18, and 22 years old
t_M	piece-wise linear decomposition with knots at 1, 30, and 70 months
r_M	piece-wise linear decomposition with knots at 30, and 70 months
t_B	piece-wise linear decomposition with knots at 1950, 1963, 1975, and 1988 years
t_C	piece-wise linear decomposition with knots at at 1, 2, and 4 cohabitations

Table C.2 – **Definitions of variables and transformations used to describe the models of the lifetime number of non-cohabiting sexual partnerships**

$\boldsymbol{Model\ LFNP1}$:
Model $LFNP1$ is a logistic model used to split individuals into two classes, either $0-10$ or $11+$ lifetime non-cohabiting sexual partnerships. Four models are fitted for four categories of A (≥ 38, $30-37$, $20-29$, and $12-19$ years old).The vector of parameters β is omitted when its role is obvious in the model formulation. The logit of the expected response for each model are:

≥ 38

$$\begin{aligned} logit(E(response)) =& r_A(A_0) \times t_M(M_0) \times M_2 \times r_M(M_1) + \\ & r_M(M_0) \times r_M(M_1) \times M_2 \times t_C(C) \times A + (r_M(M_0) + r_M(M_1)) \times t_B(B) \times A \end{aligned}$$

$30-37$

$$log(E(response)) = r_A(A_0) \times r_M(M_0) \times M_2 \times r_M(M_1) + r_M(M_0) \times r_M(M_1) \times M_2 \times t_C(C) \times A + (r_M(M_0) + r_M(M_1)) \times t_B(B) \times A$$

$20-29$

$$log(E(response)) = (r_A(A_0)) \times (r_M(M_0)) \times (M_2) \times (r_M(M_1)) + (r_M(M_0)) \times (r_M(M_1)) \times (M_2) \times (t_C(C)) \times A + (r_M(M_0) + r_M(M_1)) \times (t_B(B)) \times A$$

$12-19$

$$log(E(response)) = r_A(A_0) \times M_0 \times M_1 \times M_2 + C \times M_0 \times B \times A$$

Models $LFNPW2$, $LFNPW3$, and $LFNPW4$

Models $LFPNW2$, $LFPNW3$ and $LFPNW4$ are logistic models used to split individuals into two classes, either $0-2$ or $3-10$ for Model $LFPN2$, either $11-40$ or $41+$ for Model $LFPNW3$, either 0 or $1-2$ for Model $LFPNW4$. The logit of the expected response for these models are:

$$logit(E(response)) = r_A(A_0) \times r_M(M_0) \times M_2 \times r_M(M_1) + r_M(M_0) \times r_M(M_1) \times M_2 \times t_C(C) \times A + (r_M(M_0) + r_M(M_1)) \times t_B(B) \times A$$

Model $LFNPW5$

Model $LFPNW5$ is a logistic model used to split individuals into two classes, either 1 or 2 lifetime non-cohabiting sexual partnerships. The logit of the expected response for the model is:

$$logit(E(response)) = r_A(A_0) \times r_M(M_0) \times M_2 \times r_M(M_1) + (r_M(M_0) + r_M(M_1)) \times t_C(C) \times A + (r_M(M_0) + r_M(M_1)) \times t_B(B)$$

Models $\boldsymbol{LFNPW6}$, and $\boldsymbol{LFNPW7}$

Models $LFPNW6$ and $LFPNW7$ are truncated negative binomial regression models used to assign a number of non-cohabiting sexual partnerships between 3 and 10, and between 11 and 40 respectively. The negative binomial distributions are normalized at 0 by subtracting 3 and 11 and right-truncated at 7 and 29 respectively. The log of the expected response for the models are:

$$\begin{aligned} log(E(response)) =& r_A(A_0) \times r_M(M_0) \times M_2 \times r_M(M_1) + \\ & (M_0 + M_1) \times t_C(C) \times t_B(B) \end{aligned}$$

Model $\boldsymbol{LFNPW8}$

Model $LFPNW8$ is a power law regression model used to assign a number of non-cohabiting sexual partnerships between 41 and 1000 (around the maximum value observed for lifetime number of partners in all the databases). The model has no predictors.

Period of non-cohabiting sexual partnerships

Table C.3 gives the definitions for the variables and transformations for the model of the period of non-cohabiting partnership acquisition.

	Definitions
Variables & Classes	
N_j	number of months spent in period j
A_j	weights for non-cohabiting partnerships acquisition for every year of age
B	year of birth
NC	lifetime number of non-cohabiting partnerships
Class1	individuals with 1 non-cohabiting partner
Class2	individuals with $2-3$ non-cohabiting partner
Class3	individuals with $4-7$ non-cohabiting partner
Class4	individuals with 8 or more non-cohabiting partner
Transformations	
s_0	Class1: transformation in a single continuous variable with bounds at 4, 21, and 39 months Class2: transformation in a single continuous variable with bounds at 14, 31, and 51 months Class3: transformation in a single continuous variable with bounds at 17, 33, and 57 months Class4: transformation in a single continuous variable with bounds at 18, 41, and 72 months
t_0	Class1: transformation in a single continuous variable with bounds at 48, 135, and 226 months Class2: transformation in a single continuous variable with bounds at 23, 75, and 155 months Class3: transformation in a single continuous variable with bounds at 27, 67, and 131 months Class4: transformation in a single continuous variable with bounds at 12, 40, and 89 months
s_1	Class1: transformation in a single continuous variable with bounds at 18, and 74 months Class2: transformation in a single continuous variable with bounds at 20, and 52 months Class3: transformation in a single continuous variable with bounds at 23, and 72 months Class4: transformation in a single continuous variable with bounds at 9, and 53 months

Table C.3 – **Definitions of variables and transformations used to describe the models of the period of non-cohabiting sexual partnerships**

The models all have the form:

$$W_j = \sum_{i=1}^{N_j} exp(f_j(\beta_j,\text{predictors}))$$

f_j is a piecewise-linear function, and βs are parameters to be estimated.

For those with less than 8 non-cohabiting partnerships,

$$\begin{aligned} f_0 &= \beta_{00} + \beta_{10} s_0(N_0) \\ f_1 &= \beta_{01} + \beta_{11} t_0(N_1) \\ f_{2j} &= \beta_{02} + \beta_{12} s_1(N_{2j}) \\ f_{2j+1} &= \beta_{01} + \beta_{11} t_0(N_{2j+1}) \end{aligned}$$

For those with 8 or more non-cohabiting partnerships, f_0 is different:

$$f_0 = \beta_{00} + \beta_{10} s_0(N_0) + \beta'_{10} log(NC)$$

Model fit figures

Figures C.9-C.12 present the fit of the model to the NSFG data by birth year. In all the figures, each individual in the NSFG data was matched for age and birth year with around 200 eligible model agents. The lifetime number of partners at the maximum age (i.e., 44 or $2015 - Birthyear$) depends on the models described in C.1.3. However, the lifetime number of partners before maximum age is an extrapolation that depends on the models described in C.1.3: the NSFG data are not used to inform this extrapolation.

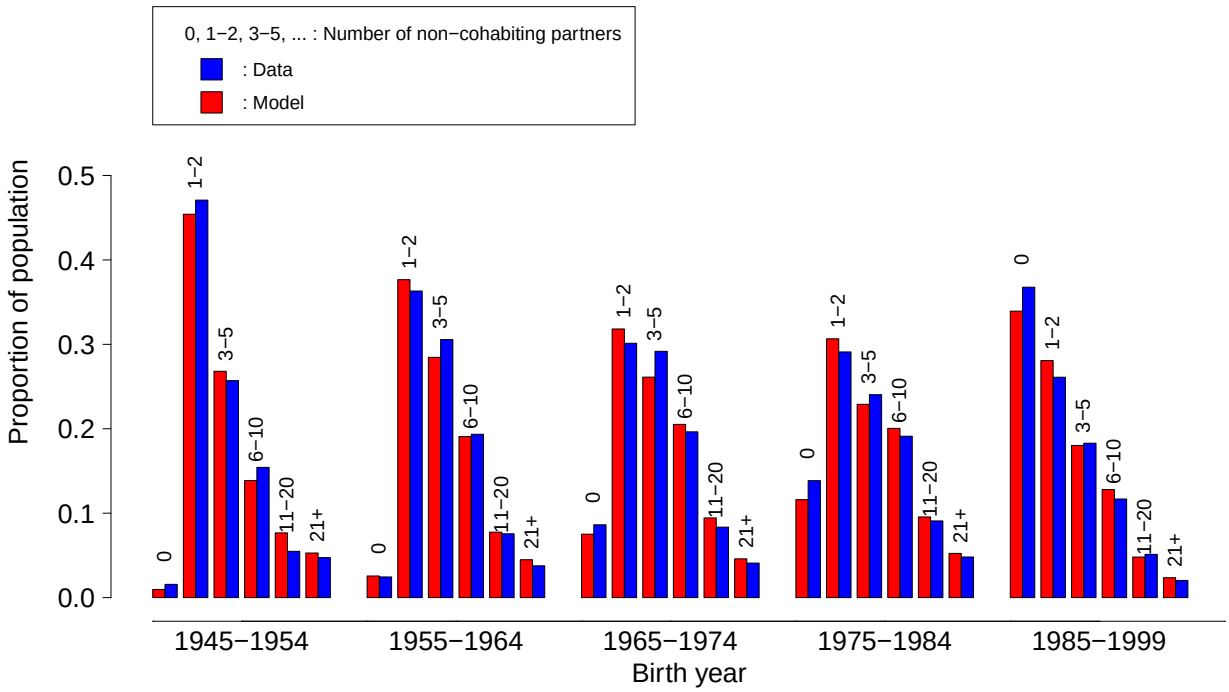


Figure C.9 – **Distributions of lifetime number of sexual partnerships for women by birth cohorts: model simulations vs NSFG data**. 21+ means 21 or more.

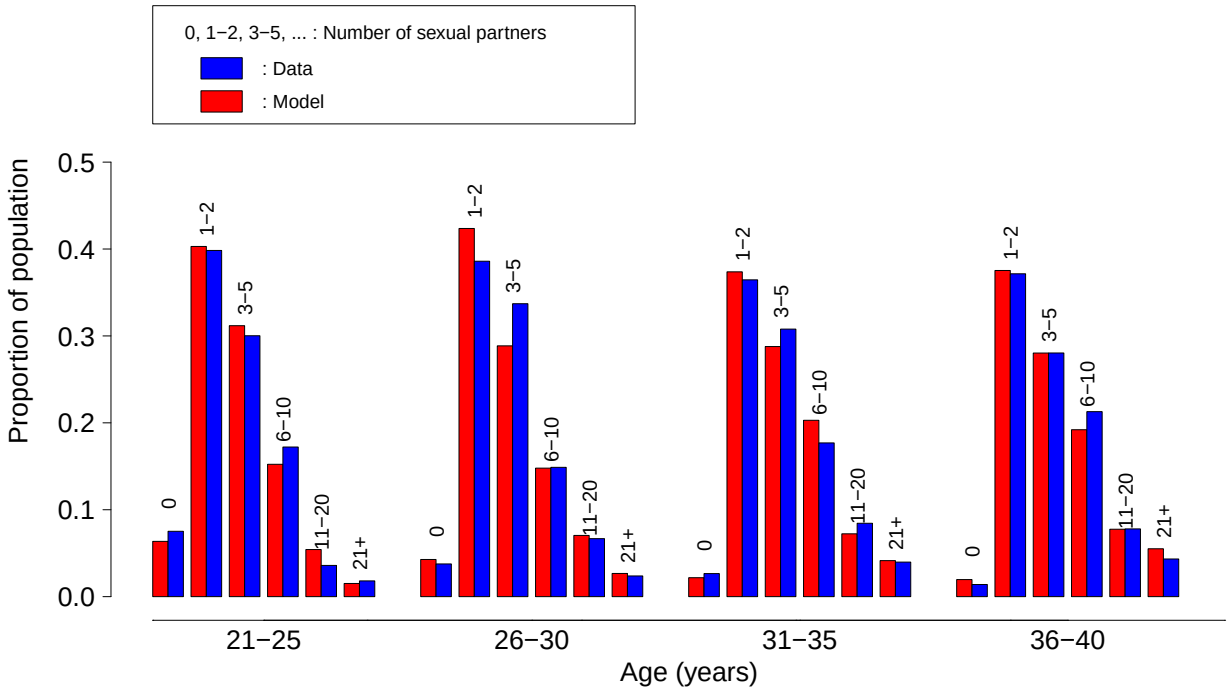


Figure C.10 – **Distributions of lifetime number of sexual partnerships by age for women born between 1955 and 1964: model simulations vs NSFG data**. 21+ means 21 or more.

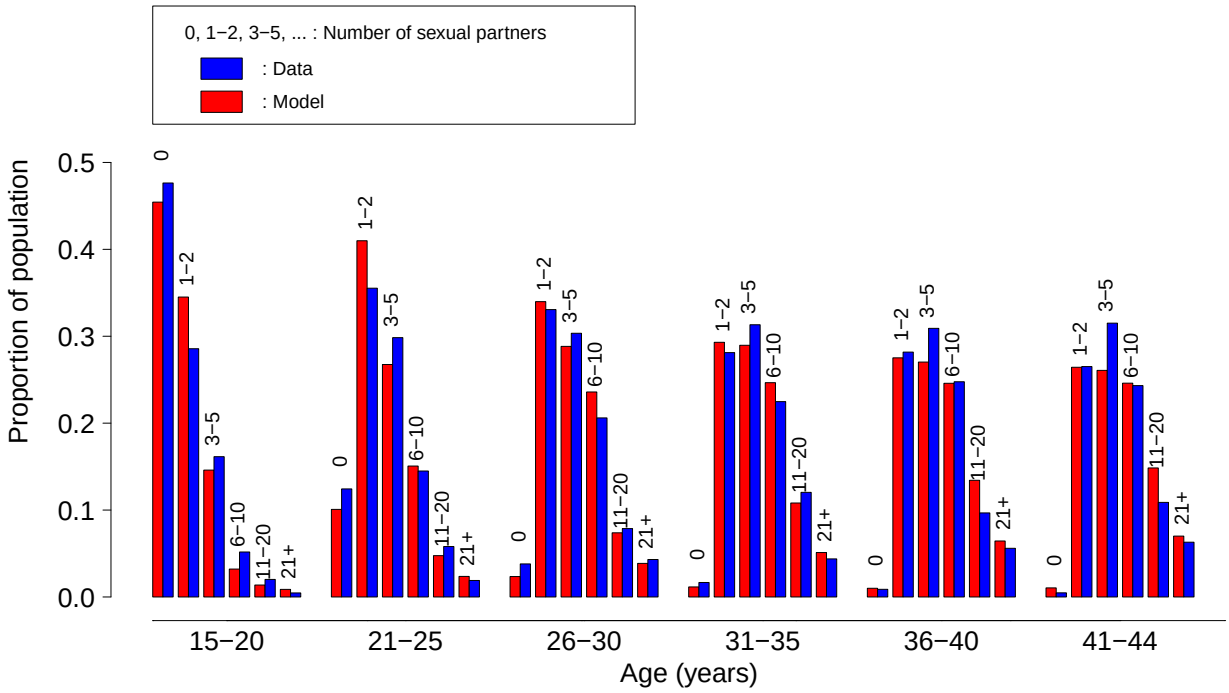


Figure C.11 – **Distributions of lifetime number of sexual partnerships by age for women born between 1965 and 1974: model simulations vs NSFG data**. 21+ means 21 or more.

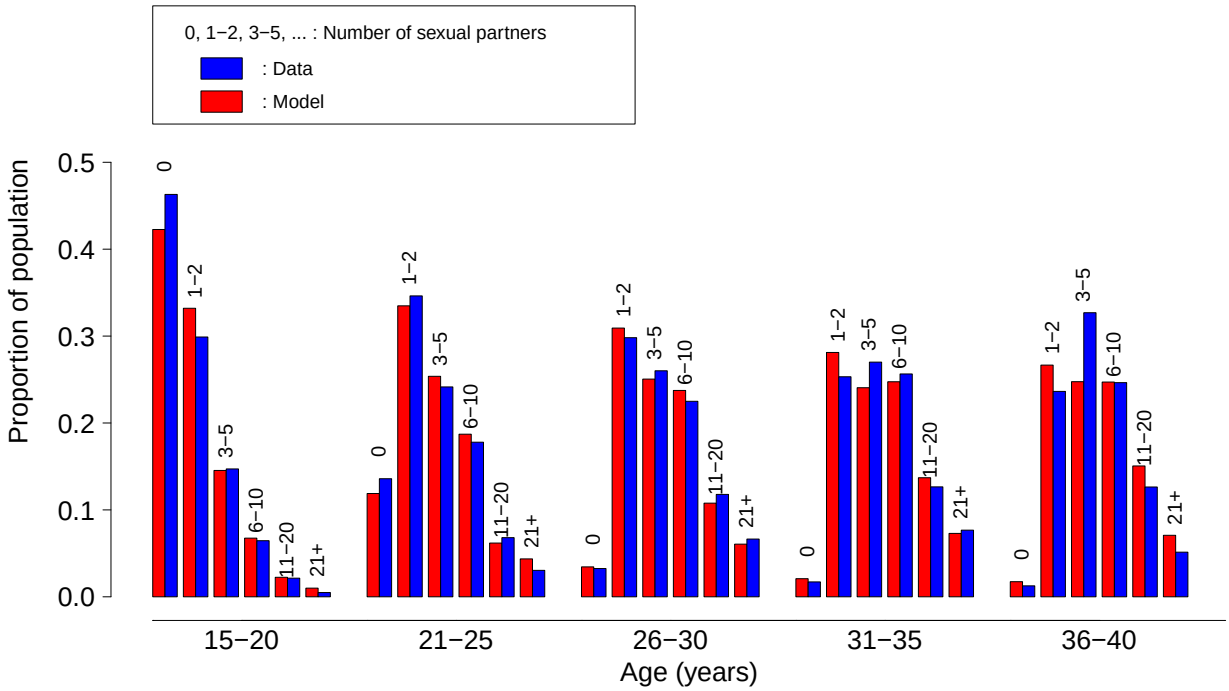


Figure C.12 – **Distributions of lifetime number of sexual partnerships by age for women born between 1975 and 1984: model simulations vs NSFG data**. 21+ means 21 or more.

C.2 Women's sexual mixing with men for birth cohorts born between 1945 and 1999

C.2.1 Cohabitation partners

Race of male partner

	Definitions
Variables	
A_w	age of wife at start of cohabitation
B	year of birth
L	length of cohabitation
C	lifetime number of cohabitations
LP	lifetime number of sex partners
Transformations	
t_B	piece-wise linear decomposition with knots at 1958, and 1975 years
t_L	categorization as: ongoing cohabitation of < 8 years, cohabitation that ended after $[0,2]$, $]2,5]$, $[5,8[$ years, cohabitation ongoing or not of duration > 8 years
t_{LP}	categorization as $1-2$, $3-10$, $11-40$, and $41+$ partners
t_{A_w}	piece-wise linear decomposition with knots at 20, 25, and 35 years

Table C.4 – **Definitions of variables and transformations used to describe the models of the race of the male cohabiting partner**

$\boldsymbol{Models\ RCPW1 - RCPW3}$

Models $RCPW1 - RCPW3$ are logistic models used to determine if the race of the partner is White or not, Black or not given that the partner is not White, Hispanic or not given that the partner is neither Black nor White respectively. The logit of the expected response for the model is:

$$logit(E(response)) = t_B(B) \times t_{LP}(LP) \times t_{A_w}(A_w) + C \times t_L(L)$$

Age of male partner

	Definitions
Variables	
A_w	age of woman at start of cohabitation
A_m	age of man at start of cohabitation
B	year of birth
L	length of cohabitation
C	lifetime number of cohabitations
LP	lifetime number of sex partners
Transformations	
t_B	piece-wise linear decomposition with knots at 1950, 1963, 1975, and 1988 years
	categorization as: ongoing cohabitation of < 8 years,
t_L	cohabitation that ended after $[0,2]$, $]2,5]$, $[5,8[$ years,
	cohabitation ongoing or not of duration > 8 years
t_{LP}	piece-wise linear decomposition with knots at 2, 5, 10, and 20 partners
t_{A_w}	piece-wise linear decomposition with knots at 20, 25, and 35 years

Table C.5 – **Definitions of variables and transformations used to describe the models of the age of the male cohabiting partner**

$\boldsymbol{Model\ ACPW1}$

Model $ACPW1$ is a logistic regression model used to determine if the value of $A_m - A_w$ is higher than 25. The logit of the expected response for the model is:

$$logit(E(response)) = t_{A_w}(A_w) \times t_B(B) + t_L(L) + C \times t_{LP}(LP) \times t_B(B)$$

$\boldsymbol{Model\ ACPW2}$

Model $ACPW2$ is a gamma regression model used to determine the value of $A_m - A_w$ if it is higher than 25. The response is normalized to 0 by subtracting 25. The log of the expected response for the model is:

$$log(E(response)) = t_{A_w}(A_w) \times t_B(B) + t_L(L) + C \times t_{LP}(LP) \times t_B(B)$$

$\boldsymbol{Model\ ACPW3}$

Model $ACPW3$ is a regression of a mixture of two normal laws used to determine the value of $A_m - A_w$ if it is lower than 25 years. Four models are fitted for four categories of A_w (≥ 35, $27-34$, $20-26$, $12-19$ years old). The log of the expected response for the normal regression model is:

$$log(E(response)) = t_{A_w}(A_w) + t_B(B)$$

The logit of the expected response for the mixture logistic model is:

$$logit(E(response)) = t_B(B) \times C + t_L(L) + t_{LP}(LP) \times B$$

Model fit figures

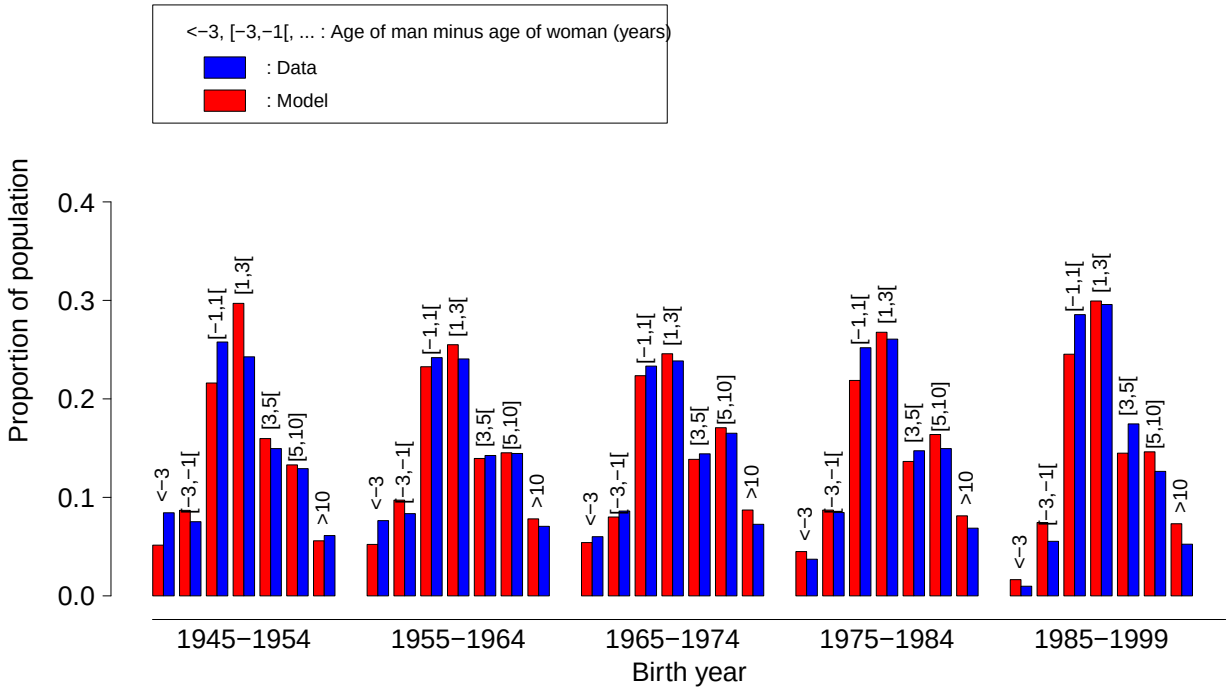


Figure C.13 – **Distributions of the difference between the age of male partner and the age of the female partner for women in cohabitations by birth year: model simulations vs NSFG data**. Women from NSFG were matched for age at start of cohabitation and birth year to women from the model.

	Definitions
Variables	
A_w	age of woman at start of partnership
A_m	age of man at start of partnership
B	year of birth
A_0	age at first sex
A_{m0}	age of male partner at woman's first sex
A_1	age at interview
C	lifetime number of cohabitations
LP	lifetime number of sex partners
Transformations	
t_A	piece-wise linear decomposition with knots at -1, 1, 3, and 8 years
	categorization as: ongoing cohabitation of < 8 years,
t_L	cohabitation that ended after [0,2],]2,5], [5,8[years,
	cohabitation ongoing or not of duration > 8 years
t_{LP}	piece-wise linear decomposition with knots at 2, 10, and 40 partners
t_B	piece-wise linear decomposition with knots at 1950, 1963, 1975, and 1988 years
s_A	piece-wise linear decomposition with knots at 16, and 20 years
r_A	piece-wise linear decomposition with knots at 16, 25, and 35 years

Table C.6 – **Definitions of variables and transformations used to describe the models of the age of the male non-cohabiting partner**

C.2.2 Non-cohabiting partners

Age of male partner

$\boldsymbol{Model\ APW1}$

Model $APW1$ is a regression of a mixture of two normal laws used to determine the value of $A_{m0} - A_0$. Four models are fitted for four categories of A_0 ($\geq$ 38, 30 $-$ 37, 20 $-$ 29, and 12 $-$ 19 years old). The log of the expected response for the mean of the normal laws is:

$$log(E(response)) = s_A(A_0)$$

The logit of the expected response for the mixture logistic model is:

$$logit(E(response)) = t_B(B) \times A_0 + t_{LP}(LP) \times A_1 \times C$$

$\boldsymbol{Model\ APW2}$

Model $APW2$ is a logistic regression model used to determine if the absolute value of $A_m - A_w$ is higher than 15. The logit of the expected response for the model is:

$$logit(E(response)) = r_{A_w}(A_w) + r_{A_1}(A_1) \times (A_1 - A_w) \times C \times t_{LP}(LP) + t_A(A_{m0} - A_0) \times s_A(A_0)$$

Model APW3

Model $APW3$ is a logistic regression model used to determine if the value of $A_m - A_w$ is higher than 0 conditional on the absolute value being higher than 15. The logit of the expected response for the model is:

$$logit(E(response)) = A_w$$

Model APW4 and *APW5*

Model $APW4$ and $APW5$ are two gamma regression models used to determine the value of $A_m - A_w$ if the absolute value is higher than 15. Model $APW4$ and $APW5$ are for positive and negative values of $A_m - A_w$ respectively. The response is normalized to 0 by multiplying by -1 for negative values and by subtracting 15. The log of the expected response for the model is:

$$log(E(response)) = A_w$$

Model APW6

Model $APW6$ is a regression of a mixture of two normal laws used to determine the value of $A_m, -A_w$ if the absolute value is lower than 15 years. Three models are fitted for three categories of A_w (≥ 35, $25 - 35$, and $12 - 24$ years old). The log of the expected response for the mean of the normal laws is:

$$log(E(response)) = r_A(A_w)$$

The logit of the expected response for the mixture logistic model is:

$$logit(E(response)) = A_1 \times (A_1 - A_w) \times t_{LP}(LP) \times C + t_A(A_{m0} - A_0) \times s_A(A_0)$$

Model fit figures

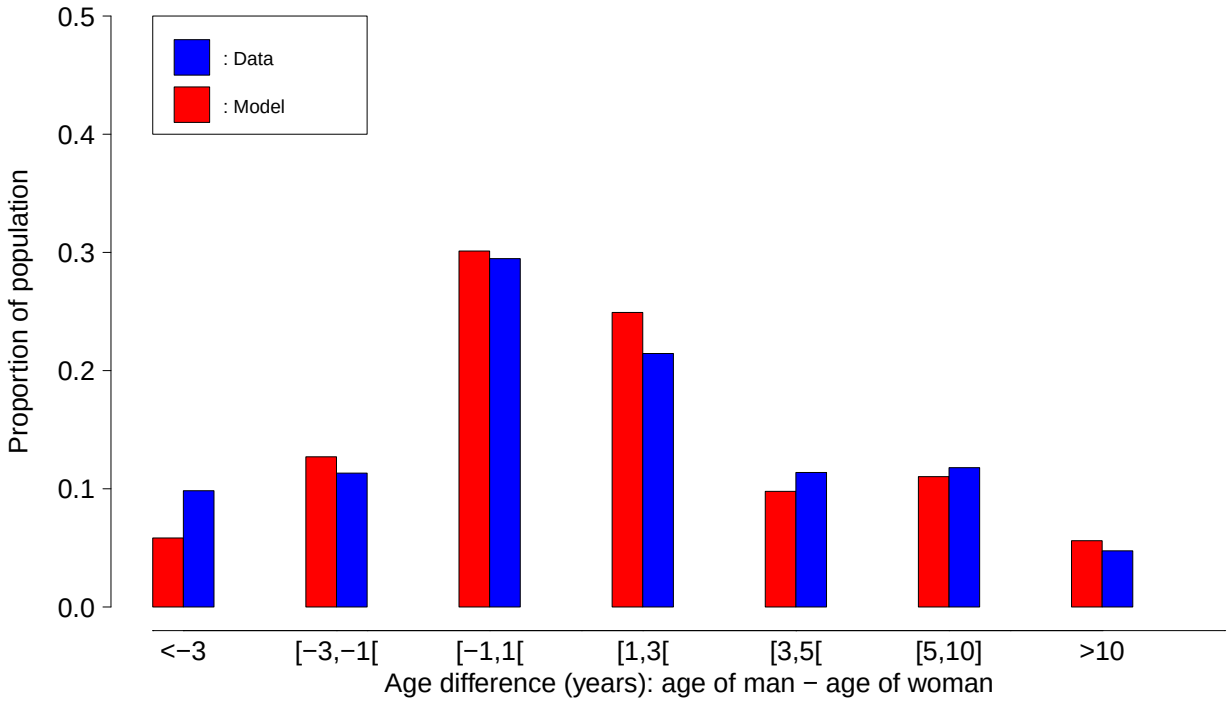


Figure C.14 – **Distributions of the difference between the age of male partner and the age of the female partner for women for non-cohabitation partnerships: model simulations vs NSFG data**. Women from NSFG were matched for age at start of partnership, birth year, and lifetime number of sexual partners to women from the model.

C.3 Incomplete profile of men's sexual history for birth cohorts born between 1945 and 1999

C.3.1 Number of cohabitations

	Definitions
Variables	
A	current age
A_0	age at first intercourse
M_0	number of months between first intercourse and first cohabitation
M_1	number of months spent outside a cohabitation after the first cohabitation
M_2	number of months spent outside a cohabitation after the first intercourse weighted proportionally by age (a month at 20 years old is worth twice a month at 40 years old)
C	lifetime number of cohabitations
B	year of birth
Transformations	
t_A	piece-wise linear decomposition with knots at 20, 25, 30, and 38 years old
r_A	piece-wise linear decomposition with knots at 14, 16, 18, and 22 years old
t_M	piece-wise linear decomposition with knots at 1, 30, and 70 months
r_M	piece-wise linear decomposition with knots at 30, and 70 months
t_B	piece-wise linear decomposition with knots at 1950, 1963, 1975, and 1988 years
t_C	piece-wise linear decomposition with knots at at 1, 2, and 4 cohabitations

Table C.7 – **Definitions of variables and transformations used to describe the models of the lifetime number of cohabitations**

The log of the response for the logistic and negative binomial models were all given by:

$$log(E(response)) = g_A(A) \times g_M(B)$$

where A is current age, B is year of birth, g_A is a piece-wise linear decomposition with knots at 20, 30, and 40, g_M is a piece-wise linear decomposition with knots at 1950, 1963, 1975, and 1988 years.

C.3.2 Number of non-cohabiting sexual partnerships

	Definitions
Variables	
A	current age
B	year of birth
C	lifetime number of cohabitations
L	length of each cohabitation
AC	age at first cohabitation
Transformations	
t_A	piece-wise linear decomposition with knots at 18, 20, 30, and 40 years old
t_B	piece-wise linear decomposition with knots at 1950, 1975, and 1988 years
t_C	categorical decomposition as 0, 1, 2, >3 cohabitations
	mean length of cohabitations
t_L	with each length transformed in a single continuous variable
	with bounds at 0.5, 1, 5, 10, 20
t_{AC}	transformation in a single continuous variable with bounds at 0.5, 1, 5, 10, and 20
	maximum length of cohabitations
s_L	with each length transformed in a single continuous variable
	with bounds at 20, 25, and 35
s_A	piece-wise linear decomposition with knots at 20, 30, and 40 years old

Table C.8 – **Definitions of variables and transformations used to describe the models of the lifetime number of non-cohabiting sexual partnerships**

$\boldsymbol{Model\ LFNPM1}$

Model $LFNPM1$ is a logistic model used to split individuals into two classes, either 0 or 1+ lifetime non-cohabiting sexual partnerships. Data for this model includes only men who have never cohabited. The logit of the expected response for the model is:

$$logit(E(response)) = t_A(A) \times t_B(B)$$

$\boldsymbol{Model\ LFNPM2}$

Model $LFNPM2$ are two logistic models used to split individuals into two classes, either 0 or 1+ lifetime non-cohabiting sexual partnerships. Data for the models includes only men who have cohabited. The two models are for men ≤ 35 and > 35 years old. The logit of the expected response for the model is:

$$logit(E(response)) = t_A(A) \times t_B(B) \times t_C(C) \times t_L(L)$$

Model LFNPM3
Model $LFNPM3$ is a logistic model used to split individuals into two classes, either $1-40$ or $41+$ lifetime non-cohabiting sexual partnerships. The logit of the expected response for the model is:

$$logit(E(response)) = s_A(A) \times t_B(B) \times (t_L(L) + s_L(L)) + t_{AC}(AC) \times B \times C \times A$$

Model LFNPM4
Model $LFNPM4$ is a logistic model used to split individuals into two classes, either $1-10$ or $11-40$ lifetime non-cohabiting sexual partnerships. The logit of the expected response for the model is:

$$logit(E(response)) = s_A(A) \times t_B(B) \times t_C(C) \times t_L(L)$$

Model LFNPM5
Model $LFNPM5$ is a logistic model used to split individuals into two classes, either $1-3$ or $4-10$ lifetime non-cohabiting sexual partnerships. The logit of the expected response for the model is:

$$logit(E(response)) = s_A(A) \times t_B(B) \times (t_L(L) + s_L(L)) + t_{AC}(AC) \times B \times C \times A$$

Models LFNPM6 − LFNMP7
ModelS $LFNPM6-LFNMP7$ are zero-inflated negative binomial models used to assign a number of non-cohabiting sexual partnerships between 1 and 3, and between 4 and 10 respectively. The distributions are normalized at 0 by subtracting 1 and 4 respectively.

$$log(E(response)) = s_A(A) \times t_C(C) + t_B(B) \times t_C(C)$$

Model LFNPM8
Model $LFNPM8$ is a logistic model used to split individuals into two classes, either $11-20$ or $21-40$ lifetime non-cohabiting sexual partnerships. The logit of the expected response for the model is:

$$logit(E(response)) = s_A(A) \times t_B(B) \times t_C(C)$$

$Models\ LFNPM9 - LFNPM10$

Model $LFNPM9 - LFNPM10$ are negative binomial models used to assign a number of non-cohabiting sexual partnerships between 11 and 20, and between 21 and 40 respectively. The negative binomial distributions are normalized at 0 by subtracting 11 and 21 respectively.

$$log(E(response)) = s_A(A) \times C + t_B(B) \times C$$

$Model\ LFNPM11$

Model $LFNPM11$ is a power law regression model used to assign a number of non-cohabiting sexual partnerships between 41 and 1000 (around the maximum value observed for lifetime number of partners in all the databases). The model has no predictors.

Correction for men born between 1945 and 1960

In the NHSLS data, the question asked to participants was:

"Now thinking about the time since your 18th birthday (again,including the recent past that you have already told us about) how many female partners have you ever had sex with?"

There is also a question on partners before the 18th birthday. In both cases, the question doesn't refer to intercourse specifically. The same question is asked in the GSS surveys (they have a common methodology having both been done by the National Opinion Research Center (NORC)). Hence, there may be a problem in comparability between NSFG and NHSLS/GSS. Comparing the distribution of lifetime number of partners for NHSLS (including sex partners before 18 years old) and NSFG for the cohort born between 1957 and 1962 (aged around 32 years in NHSLS and 42 years in NSFG, but this is adjusted for in the figure) in Figure C.15 we see that the tail of distribution is heavier in NHSLS. We observe the same when comparing GSS with NSFG. This would make sense if people with many intercourse partners also happen to have many non-intercourse partners such as oral sex partners. To correct for that, we used data from NSFG on sex partners that include oral and anal sex partners asked in the Audio Computer-Assisted Self-Interviewing (ACASI) section to fit two models. The first model estimates the probability of having a higher number of partners declared in the ACASI question compared to the Computer-Assisted Personal Interview (CAPI) question on intercourse only, and the second model estimates the difference in the number of partners between the two answers. We thus used these models to estimate the number of non-intercourse partners for participants in the NHSLS.

$$logit(P(D > 0)) = g_{LP}(LP) \times g_A(A)$$

where D is the difference between the ACASI and CAPI responses. LP is the lifetime number of partners in the ACASI question, g_{LP} is a piece-wise linear decomposition with knots at 2, 5, 10, and 40. A is the age at interview and g_A is a piece-wise linear decomposition with knots at 20, 30, and 40 years.

$$log(E(D-1|D>0)) = g_{LP}(LP) \times g_A(A)$$

where $D-1$ is assumed to have a negative binomial distribution if $D > 0$.
As Figure C.15 shows, the correction improves the fit, but the tail of the distribution for the NHSLS data remains heavier. Hence, we cannot exclude the possibility of an overestimation of men with high number of partners (> 20) for the older birth cohorts. We compare in Figure C.16 the final lifetime number of partners of the men in the model with the data from the GSS on the number of sex partners after the 18th birthday to verify that the trends are preserved.

Model fit figures

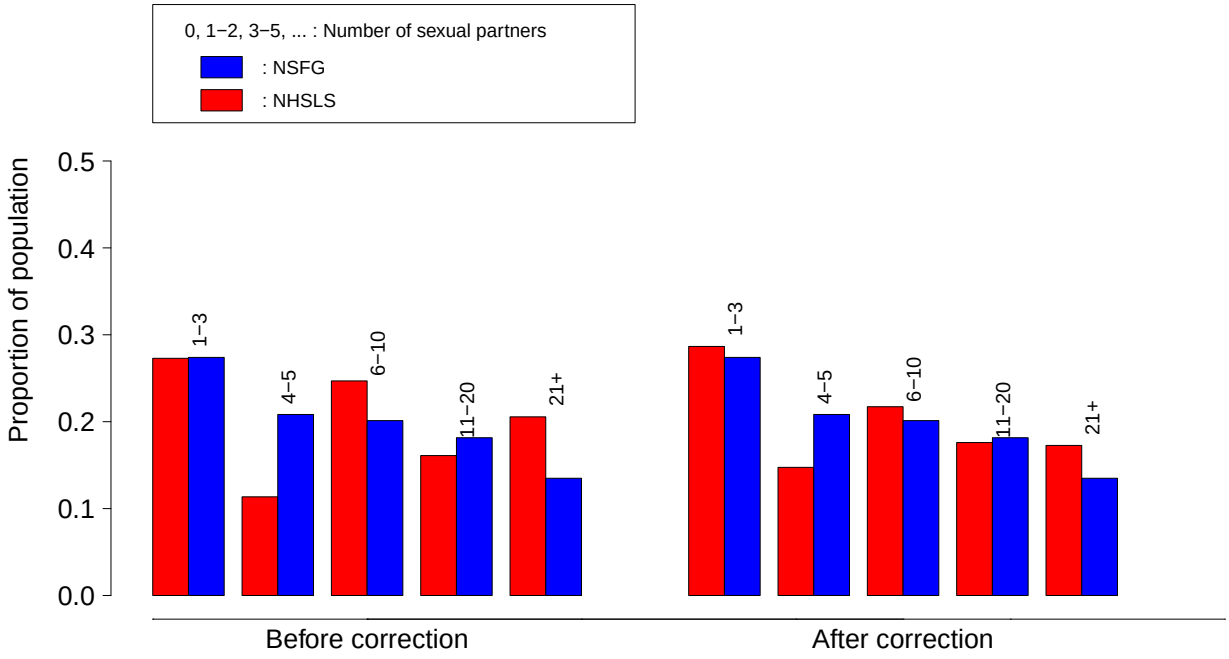


Figure C.15 – **Distributions of lifetime number of sexual partnerships for men born between 1957 and 1962: NHSLS data vs NSFG data**. 21+ means 21 or more. Men from NSFG were matched for age and birth year to men from NHSLS. The number of lifetime partners at younger age for the NSFG participants by using their cohabitation history and the models described in section C.7.2

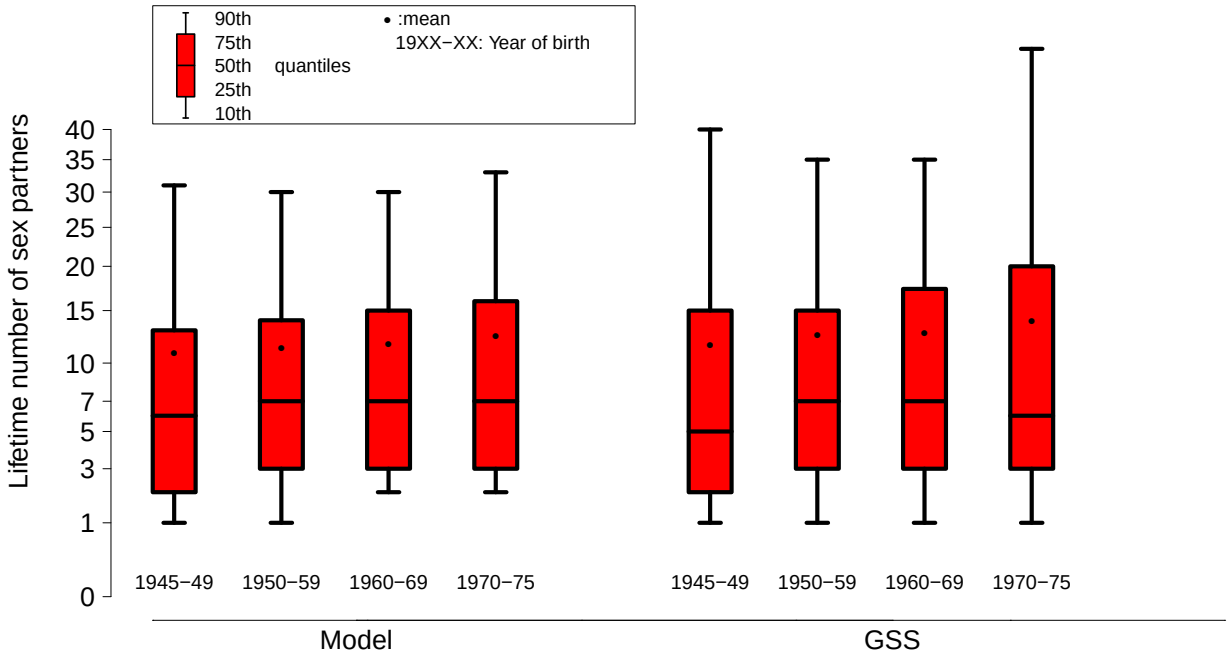


Figure C.16 – **Distributions of lifetime number of sexual partnerships for men by birth year: model simulations vs GSS data**. Men from the model are all 44 years old and sex partners before 18 years old are included. Male participants from GSS are between 40 and 50 years old and sex partners before 18 years old are not included.

C.3.3 Race of female cohabiting partner

	Definitions
Variables	
A	current age
B	year of birth
L	length of cohabitation
S	indicator for whether the cohabitation is ongoing or has ended
Transformations	
t_B	piece-wise linear decomposition with knots at 1950, 1975, and 1988 years
t_L	categorization as [0,0.5],]0.5,1],]1,5],]5,8], and 8+ years

Table C.9 – Definitions of variables and transformations used to describe the models of the race of the female cohabiting partner

$\boldsymbol{Models\ RCPM1 - RCPM3}$

Models $RCPM1 - RCPM3$ are logistic models used to determine if the race of the partner is White or not, Black or not given that the partner is not White, Hispanic or not given that the partner is neither Black nor White respectively. The logit of the expected response for the model is:

$$logit(E(response)) = S \times t_L(L) \times t_B(B)$$

C.3.4 Age of female cohabiting partner

	Definitions
Variables	
A	age of man at interview
A_m	age of man at start of cohabitation
A_w	age of woman at start of cohabitation
B	year of birth
L	length of cohabitation
S	indicator for whether the cohabitation is ongoing or has ended
LP	lifetime number of sex partners
NC	the number of cohabitations preceding the current cohabitation
Transformations	
t_B	piece-wise linear decomposition with knots at 1962, and 1975 years
t_L	transformed in a single continuous variable with bounds at 1, 5, 10, and 20 years
t_{A_m}	piece-wise linear decomposition with knots at 20, 25, and 35 years
t_{LP}	piece-wise linear decomposition with knots at 3, 10, and 20 partners
s_L	piece-wise linear decomposition with knots at 2, 3, and 4

Table C.10 – **Definitions of variables and transformations used to describe the models of the age of the female cohabiting partner**

$ModelACPM1$

Model $ACPM1$ is a regression of a mixture of two normal laws used to determine the value of $A_m - A_w$. Four models are fitted for four categories of A_m (≥ 35, $25-34$, $20-24$, and $12-19$ years old). The log of the expected response for the normal regression model is:

$$log(E(response)) = t_B(B)$$

The logit of the expected response for the mixture logistic model is:

$$log(E(response)) = A \times t_{LP}(LP) \times NC \times t_L(L) + B \times s_L(t_L(L)) \times S$$

C.4 Old Cohorts

C.4.1 Number of partnerships after 18 years old

The models are the same for both men and women, the models are the same as Model 1 and 3 to 11 from section C.3.2 except that the predictors are simply the year of birth transformed as two piece-wise linear variables with knot at 1920. Of note, in older cohorts we cannot exclude a survivor bias

(e.g., people with less sex partners living longer) or a stronger recall bias among these cohorts. Fits of models that include age as a predictor support this.

C.4.2 Age at first marriage

The multinomial model has only the number of sex partners after 18 years of age as predictor. The variable is categorized as $0-2$, $3-5$, $6-10$, $11-40$, and $41+$.

Model fit figures

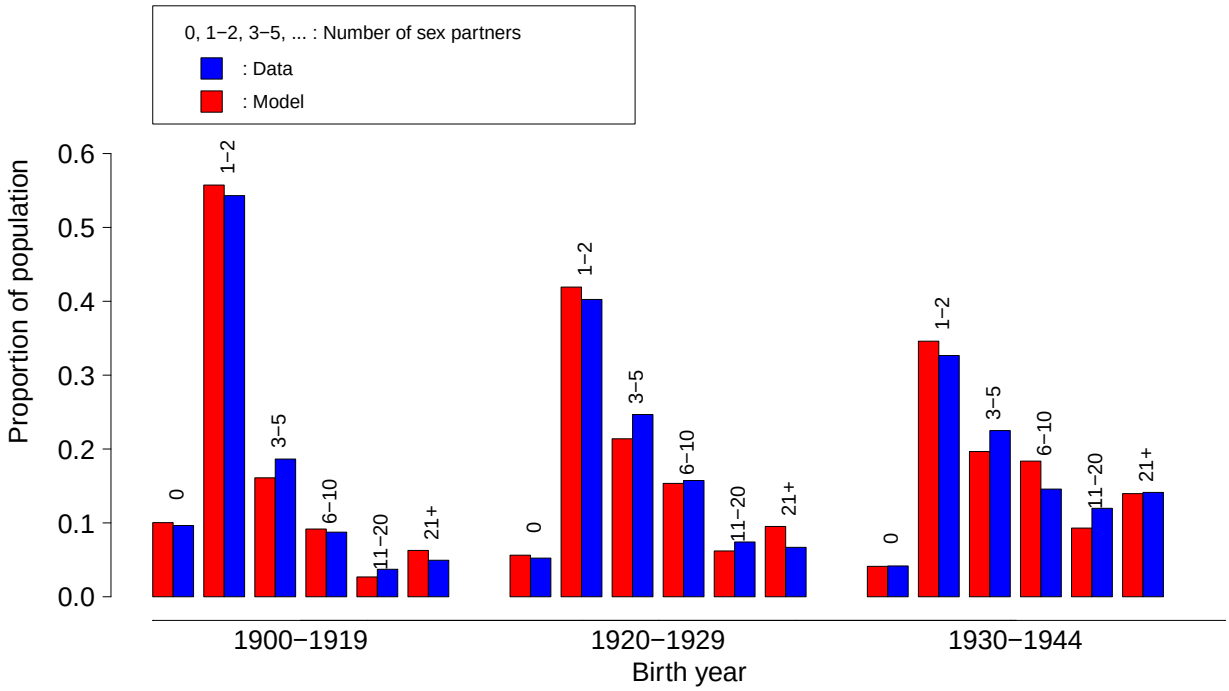


Figure C.17 – **Distributions of lifetime number of sexual partnerships for men by birth cohorts: model simulations vs GSS data**. 21+ means 21 or more. Men from GSS were matched for birth year to men in the model.

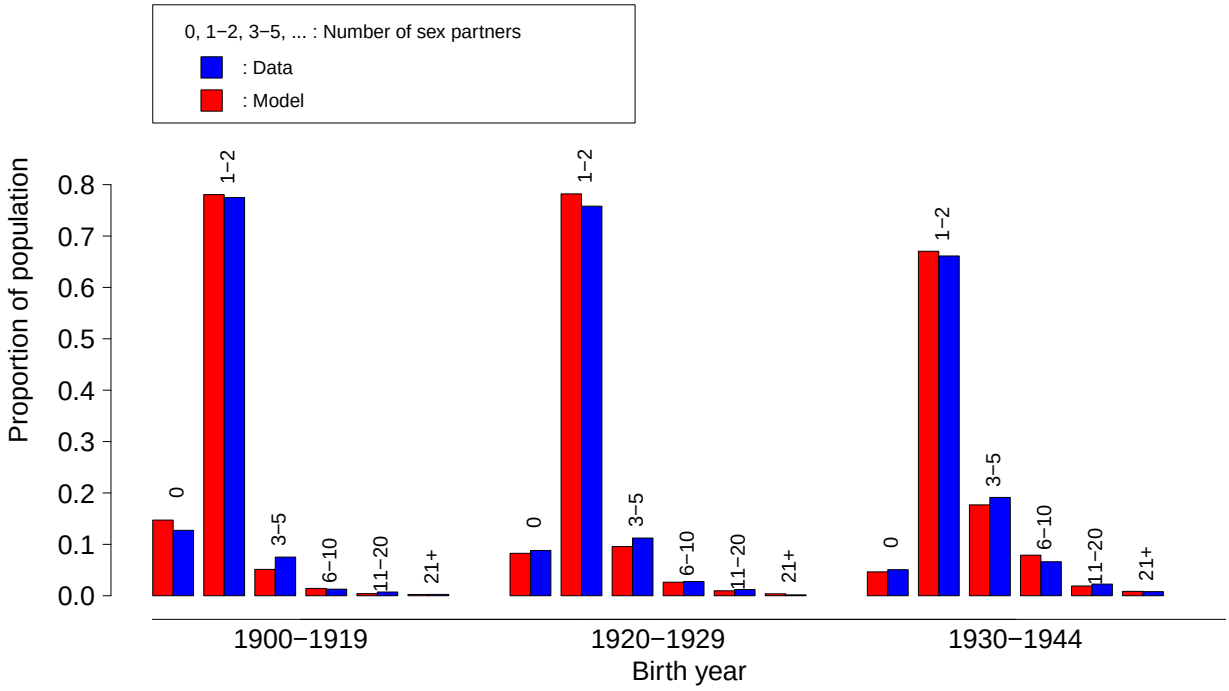


Figure C.18 – **Distributions of lifetime number of sexual partnerships for women by birth cohorts: model simulations vs GSS data**. 21+ means 21 or more. Women from GSS were matched for birth year to women in the model.

C.5 Who mixes with whom (cohabitation)

C.5.1 Lifetime number of partners of the partner

Female and male partners outside the model population are given a category of lifetime number of partners ($1-4$, $5-10$, $11-20$ and $21+$). Male partner inside the model population also have their lifetime number of partners category determined before the sampling of the male partner. To determine the category of lifetime number of partners, we used three logistic models for men and three logistic models for women.

	Definitions
Variables	
AD	age difference: age of man minus age of woman
A_s	age of participant at start of cohabitation
B	year of birth
L	length of cohabitation
LP	lifetime number of sex partners
S	indicator for whether the cohabitation is ongoing or has ended
T	time since end of last cohabitation
Transformations	
t_B	piece-wise linear decomposition with knots at 1962, and 1975 years
t_L	categorization as [0,1],]1,5], [5,10],]10,20] and 21+ years
t_{A_s}	categorization as [0,20],]20,25],]25,35], 35+ years
t_{AD}	categorization as ≤ -6, $]-6,-3]$, $]-3,-1]$, $]-1$,1],]1,3],]3,8], and >8 years
t_{LP}	categorization as $1-4$, $5-10$, $11-20$, and $21+$ partners
g_T	piece-wise linear decomposition with knots at $1/12$, 1, 3, 5, and 8 years

Table C.11 – **Definitions of variables and transformations used to describe the models of the lifetime number of partners of the cohabiting partners**

C.5.2 Models $LFNPPC1 - LFNPPC3$

Models $LFNPPC1 - LFNPPC3$ are logistic models used to determine the category of lifetime number of partners for the cohabiting partner: $t_{LP}(LP)$. The models follow a conditional structure: Model $LFNPPC1$ determines whether the partner has between 1 and 4 partners or not, Model $LFNPPC2$ determines whether the partner has between 5 and 10 partners conditional on the partner having more than 4, and Model $LFNPPC3$ determines if the partner has between 11 and 20 partners conditional on the partner having more than 10.

$$logit(E(response)) = t_{A_s}(A_s) \times t_{AD}(AD) \times t_B(B) + t_B(B) \times t_L(L) \times S$$

C.5.3 Cohabitations with individuals outside the model

Men will have female partners outside the model population. To determine if an upcoming cohabitation occurs with a female outside the model population, it is necessary to first impute the hypothetical age at the start of the next cohabitation. To do this, we fit a model:

$$log\left(\frac{1}{\lambda(predictors)}\right) = t_B(B) \times t_A(A) \times g_T(T)$$

where λ is the rate of initiation of the upcoming cohabitation, β is the vector of parameters.

C.6 Complete the profile of cohabitation history of men

C.6.1 Model fit figures

As Figure C.22 shows, age at first cohabitation for the cohort born between 1945 and 1954 is older in the model simulations compared to the NHSLS data. In the empirical data, there is a sharp increase in the age of first cohabitation between 1945 and 1965 that is reproduced to a much lesser extent in the model simulations. Qualitatively, however, the trends are preserved.

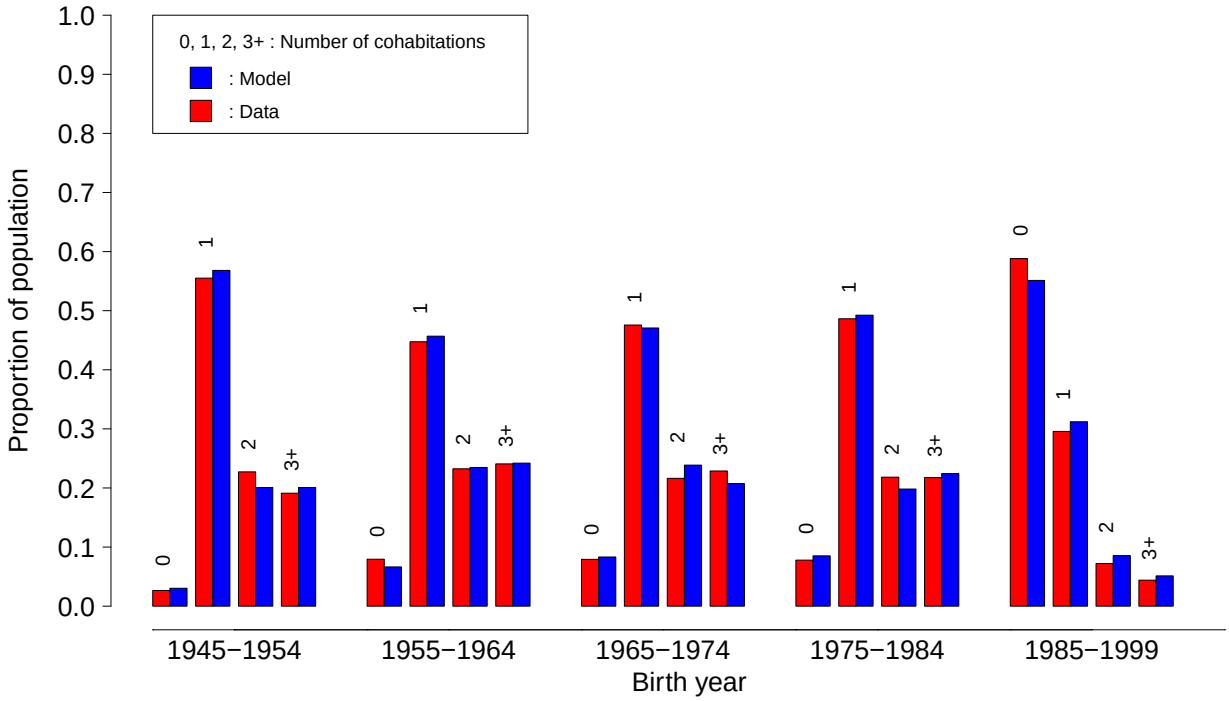


Figure C.19 – **Distributions of lifetime number of cohabitations for men by birth cohorts before mixing occurs: model simulations vs NHSLS and NSFG data**. 3+ means 3 or more. Cohabitations of men have not yet been assigned a female partner.

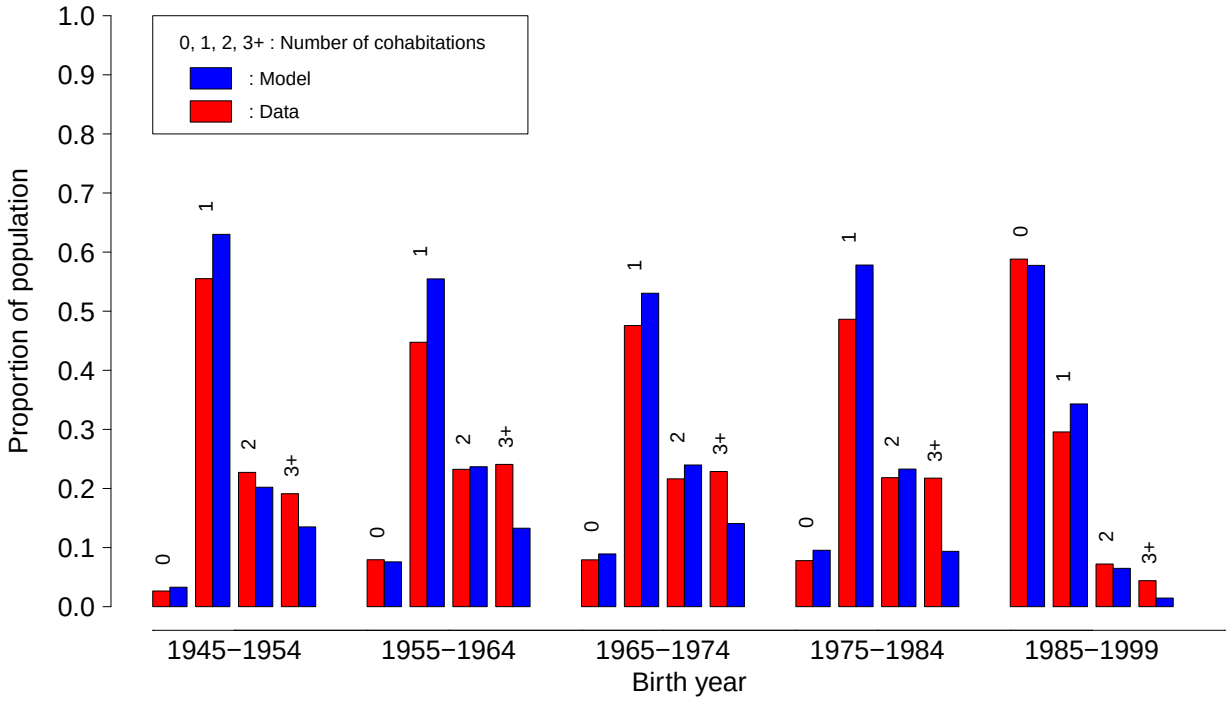


Figure C.20 – **Distributions of lifetime number of cohabitations for men by birth cohorts after mixing occurs: model simulations vs NHSLS and NSFG data**. 3+ means 3 or more. Cohabitations of men have been assigned a female partner. Cohabitations without a female partner assigned were converted to non-cohabiting partnerships.

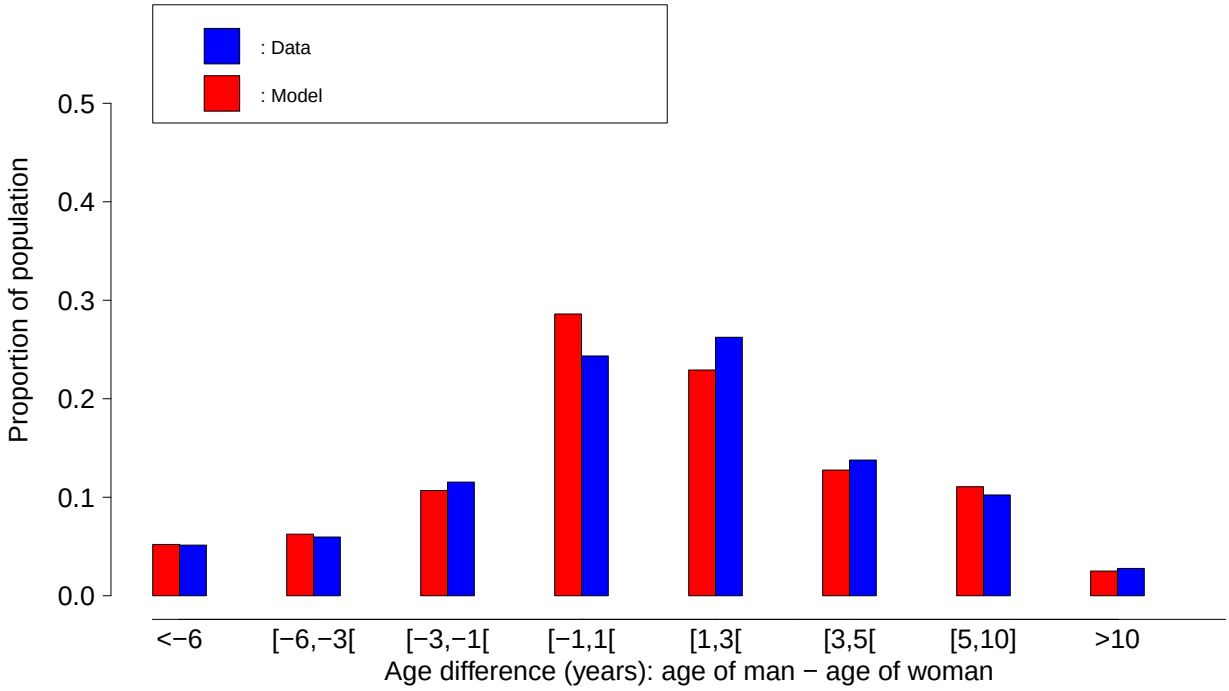


Figure C.21 – **Distributions of the difference between the age of male partner and the age of the female partner for men in cohabitations: model simulations vs NSFG/NHSLS data**. Men from NSFG and NHSLS were matched for age at start of cohabitation and birth year to men from the model.

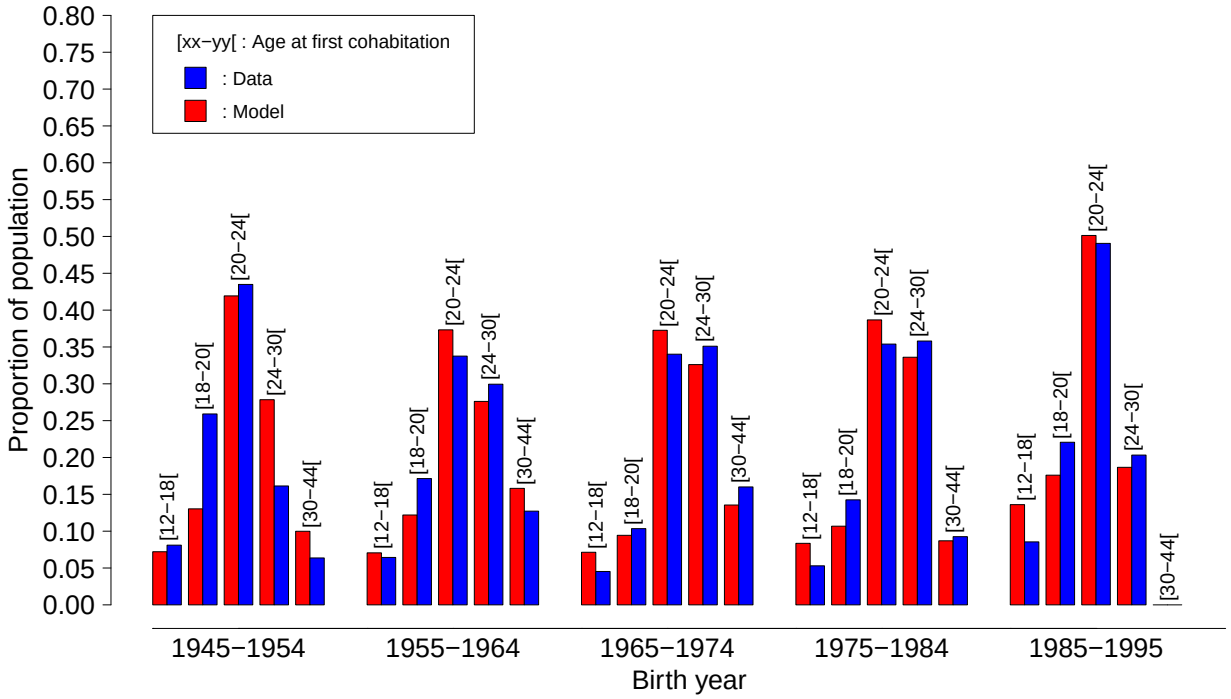


Figure C.22 – **Distributions of the age at start of first cohabitation for men by birth year: model simulations vs NSFG/NHSLS data**.

C.7 Estimate an incomplete profile of non-cohabiting partnerships

C.7.1 Estimating the age at first sex of men

	Definitions
Variables	
AC	age at start of first cohabitation
B	year of birth
A	age at interview
LP	lifetime number of sex partners
Transformations	
t_{AC}	piece-wise linear decomposition with knots at 20, 27, and 35 years
t_{LP}	piece-wise linear decomposition with knots at 5, 20, and 40 partners
t_A	piece-wise linear decomposition with knots at 25, and 35 years old

Table C.12 – **Definitions of variables and transformations used to describe the models of the age at first sex for men**

$\boldsymbol{Models\ AFSM1 - AFSM2}$

Models $AFSM1 - AFSM2$ are logistic models used to determine if the length of the gap between the date of first sex and the start of the first cohabitation is lower than 3 years and, if not, if the length of the gap is lower than 9 years. The models only apply to men who have had a cohabitation in their lifetime:

$$logit(E(response)) = (A + B) \times t_{LP}(LP) \times t_{AC}(AC)$$

$\boldsymbol{ModelsAFSM3 - AFSM5}$

Models $AFSM3 - AFSM5$ are negative binomial models used to determine the exact length of the gap in months. The models only apply to men who have had a cohabitation in their lifetime:

$$log(E(response)) = A \times t_{LP}(LP) + t_{AC}(AC)$$

$\boldsymbol{Models\ AFSM6 - AFSM7}$

Models $AFSM6 - AFSM7$ are logistic models used to determine if the age at first sex is lower or equal to 16 years old, and lower or equal to 18 years old if it is higher than 16 years old. The models only apply to men who never had a cohabitation in their lifetime:

$$logit(E(response)) = B \times t_{LP}(LP) \times t_A(A)$$

Models AFSM8 − AFSM9

Models $AFSM8 - AFSM9$ are negative binomial models used to determine the age at first sex (in years) if it is lower than 16 years old, and if it is higher than 18 years old. The models only apply to men who never had a cohabitation in their lifetime:

$$log(E(response)) = t_{LP}(LP) \times t_A(A)$$

Models AFSM10

Model $AFSM10$ is a logistic models used to determine the age at first sex (in years) if it is 17 or 18 years old. The models only apply to men who never had a cohabitation in their lifetime:

$$logit(E(response)) = t_{LP}(LP) \times t_A(A)$$

C.7.2 Estimating the period of non-cohabiting sexual partnerships for men

The models used to estimate the period are the same as for women (see section C.1.3).

C.7.3 Age of female partner

	Definitions
Variables	
A	age of man at interview
A_m	age of man at start of partnership
A_w	age of woman at start of start of partnership
LP	lifetime number of sex partners
LC	lifetime number of cohabitations
Transformations	
t_{LP}	piece-wise linear decomposition with knots at 5, 20, and 40 partners

Table C.13 – **Definitions of variables and transformations used to describe the models of the age of the female non-cohabiting partner**

$\boldsymbol{Model\ APM1}$

Model $APM1$ is a regression of a mixture of two normal laws used to determine the value of $A_m - A_w$. The model is fitted for three categories of A ($\geq$ 30, $19-29$, and $12-18$ years old). The log of the expected response for the normal regression model is:

$$log(E(response)) = A$$

The logit of the expected response for the mixture logistic model is:

$$logit(E(response)) = A \times t_{LP}(LP) \times (A - A_m) \times LC$$

C.7.4 Men who pay for sex

	Definitions
Variables	
A	age of man at interview
B	year of birth
LP	lifetime number of sex partners
Transformations	
t_{LP}	piece-wise linear decomposition with knots at 4, 20, and 40 partners
t_B	piece-wise linear decomposition with knots at 1920, 1940, 1960, and 1980 years
t_A	piece-wise linear decomposition with knots at 20, 35, and 50 years

Table C.14 – **Definitions of variables and transformations used to describe the models to identify men who have paid for sex**

Model MWPS

Models $MWPS$ is a logistic model used to determine if a man has paid for sex in his lifetime:

$$logit(E(response)) = t_A(A) \times t_{LP}(LP) \times t_B(B)$$

C.7.5 Model fit figures

We see from Figure C.23 that age at first sex for men born before 1955 is slightly older in the model simulations vs the NHSLS data. This is likely a consequence of age at start of first cohabitation not fitting well for these birth cohorts (see section C.6.1).

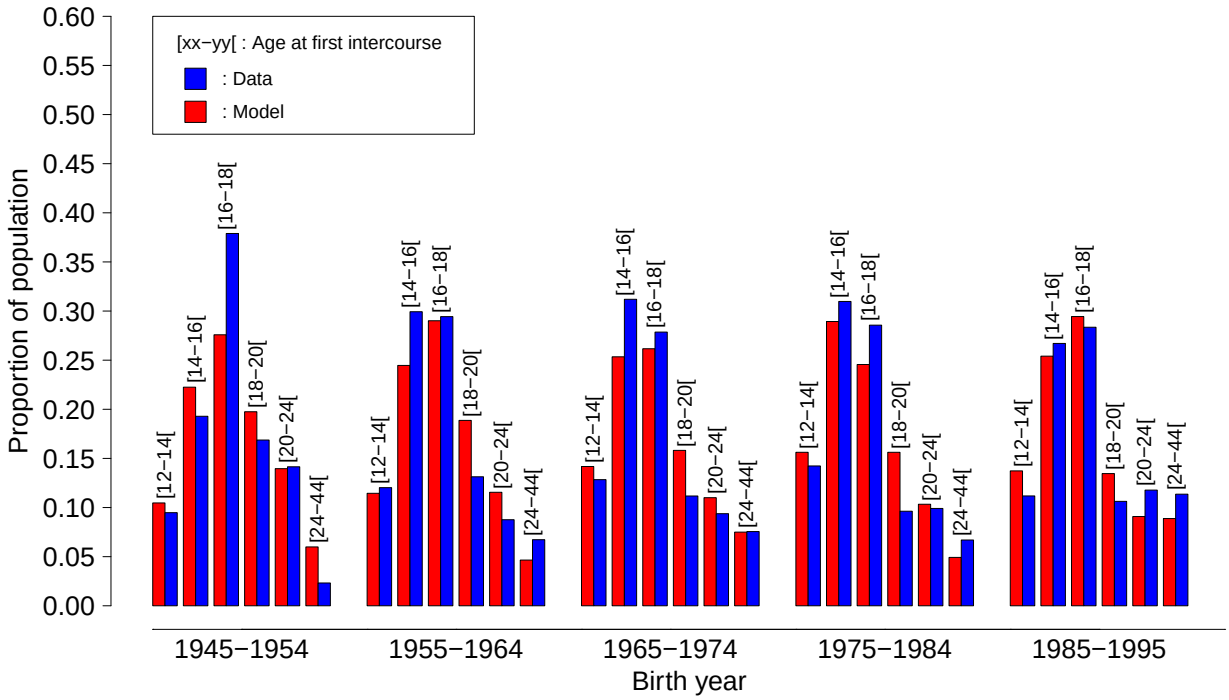


Figure C.23 – **Distributions of the age at first sex for men by birth year: model simulations vs NSFG/NHSLS data**.

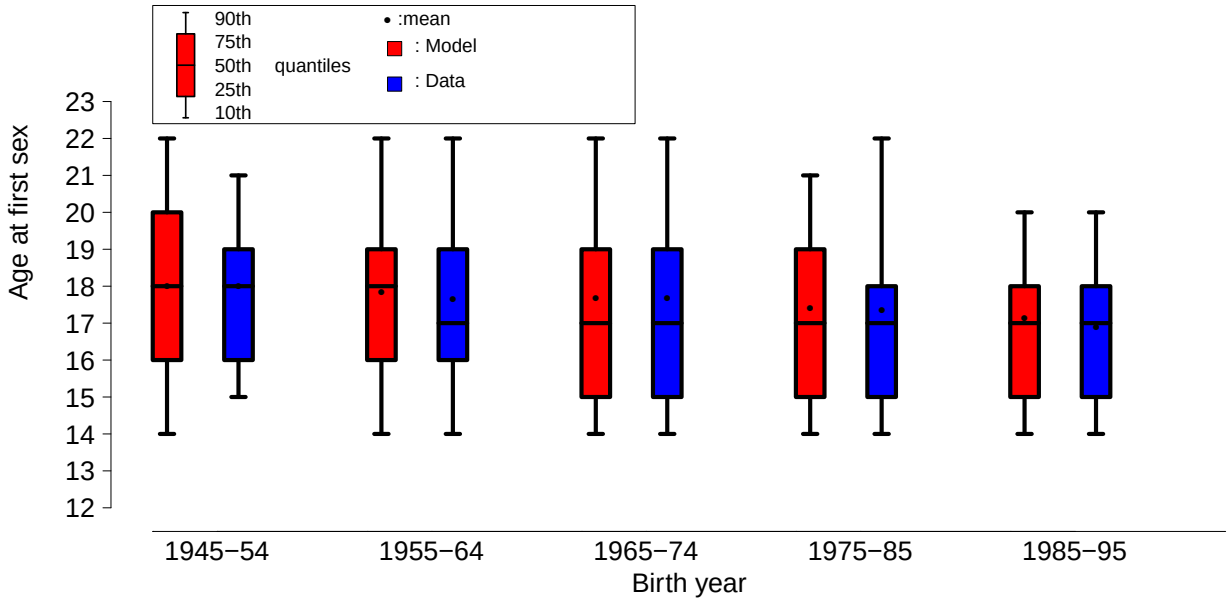


Figure C.24 – **Quantiles of the age at first sex for men by birth year: model simulations vs NSFG/NHSLS data**.

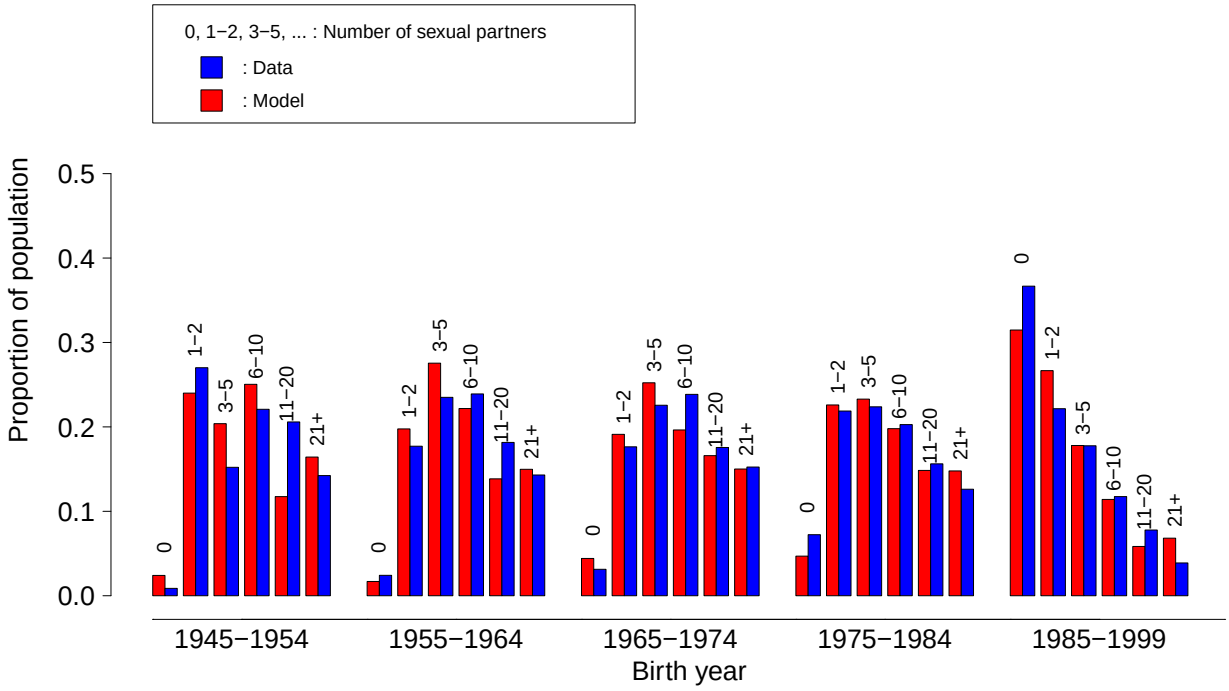


Figure C.25 – **Distributions of lifetime number of sexual partnerships for men by birth cohorts: model simulations vs NHSLS/NSFG data**. 21+ means 21 or more. Men participants from NHSLS/NSFG were matched for age and birth year with men from the model.

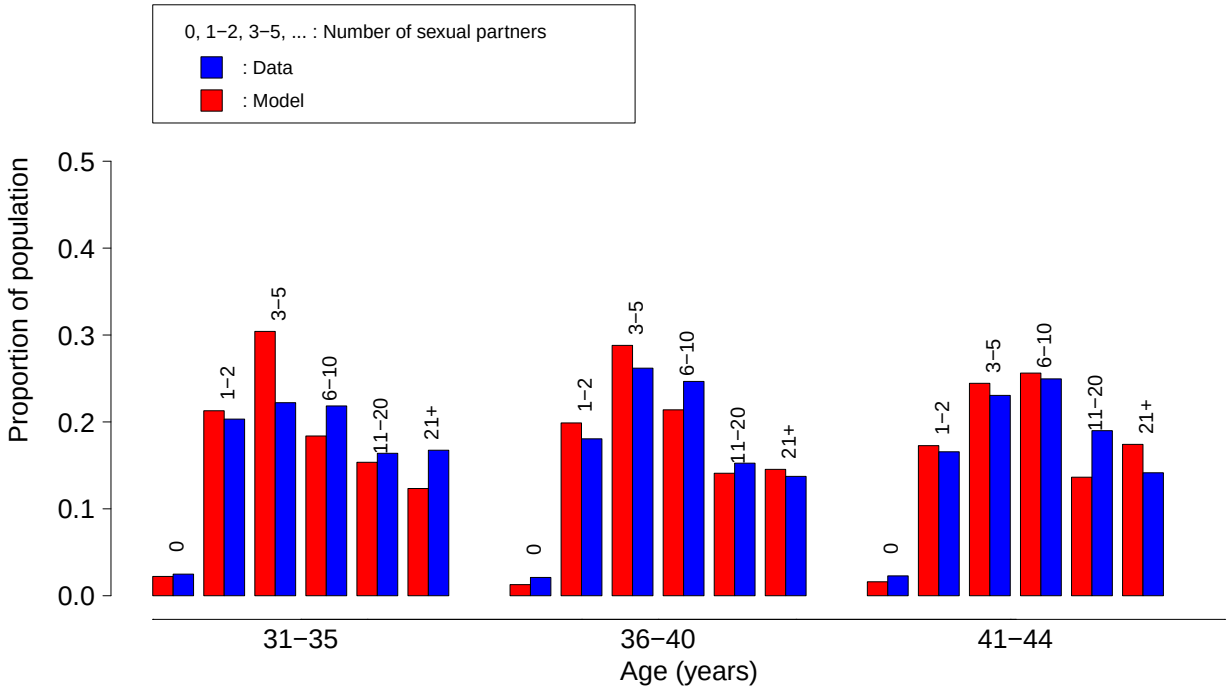


Figure C.26 – **Distributions of lifetime number of sexual partnerships by age for men born between 1955 and 1964: model simulations vs NHSLS/NSFG data**. 21+ means 21 or more. Men participants from NHSLS/NSFG were matched for age and birth year with men from the model.

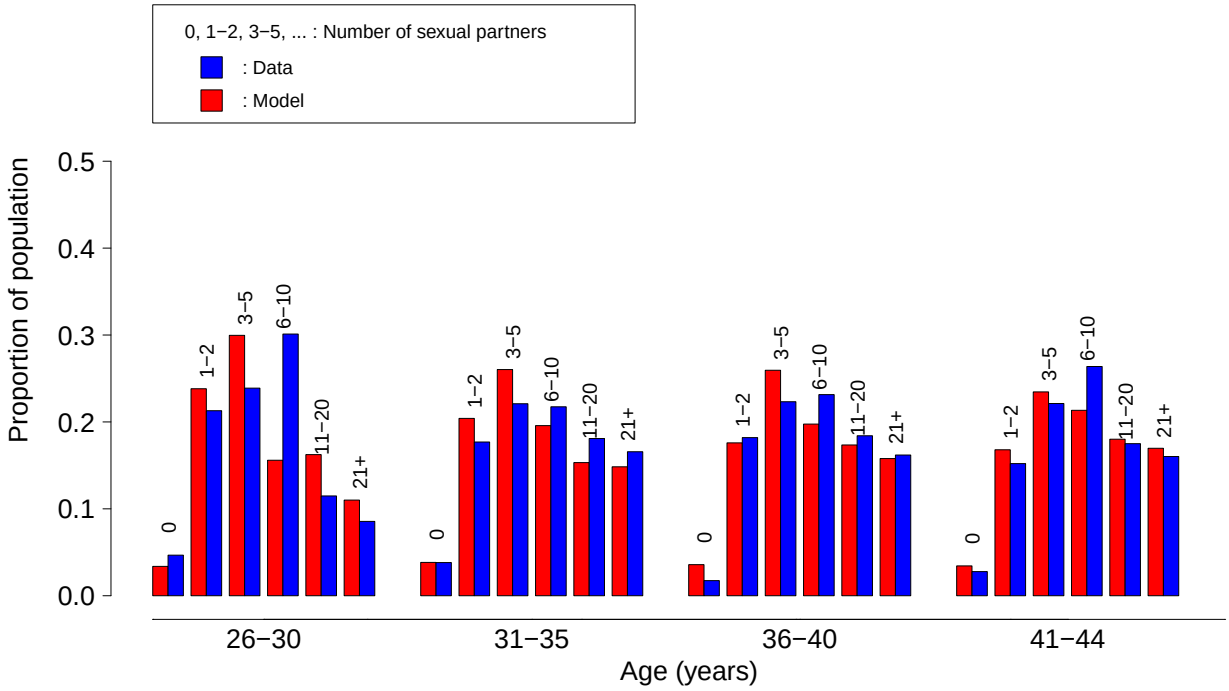


Figure C.27 – **Distributions of lifetime number of sexual partnerships by age for men born between 1965 and 1974: model simulations vs NHSLS/NSFG data**. 21+ means 21 or more. Men participants from NHSLS/NSFG were matched for age and birth year with men from the model.

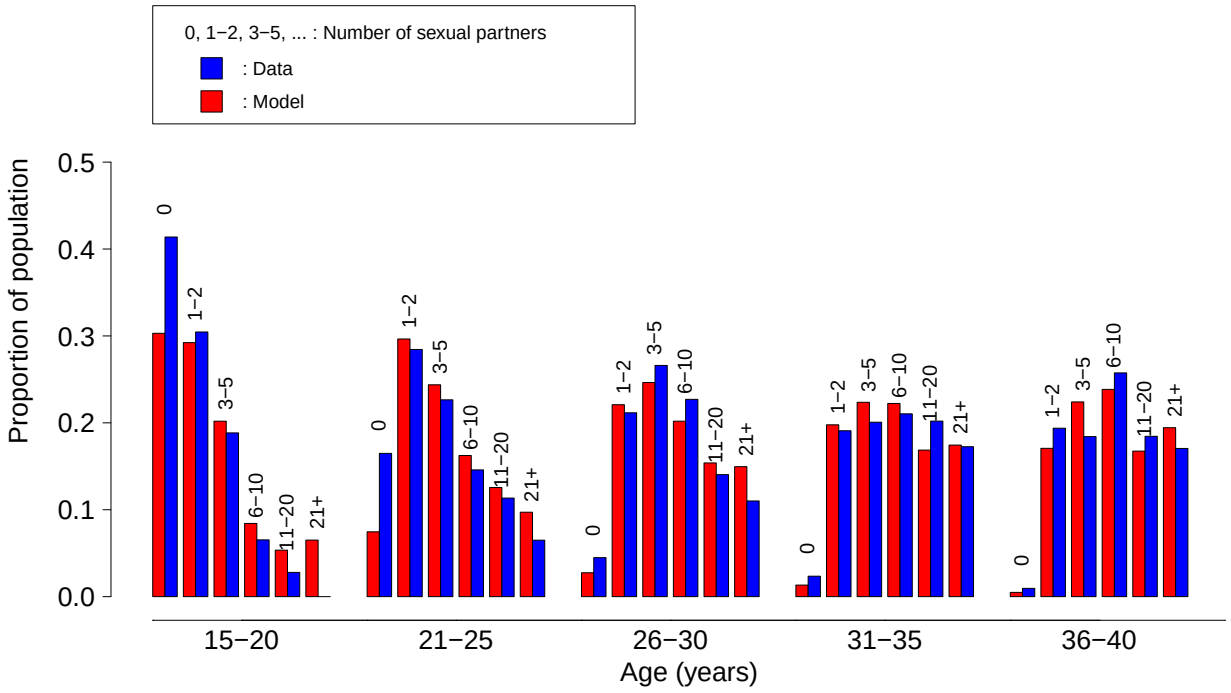


Figure C.28 – **Distributions of lifetime number of sexual partnerships by age for men born between 1975 and 1984: model simulations vs NHSLS/NSFG data**. 21+ means 21 or more. Men participants from NHSLS/NSFG were matched for age and birth year with men from the model.

Figure C.29 – **Proportion of men who have paid for sex by year of birth: model simulations**.

C.8 Balancing non-cohabiting partnerships

C.8.1 Scenario 1: Under-sampling and under-reporting of a core group of women

Model fit figures

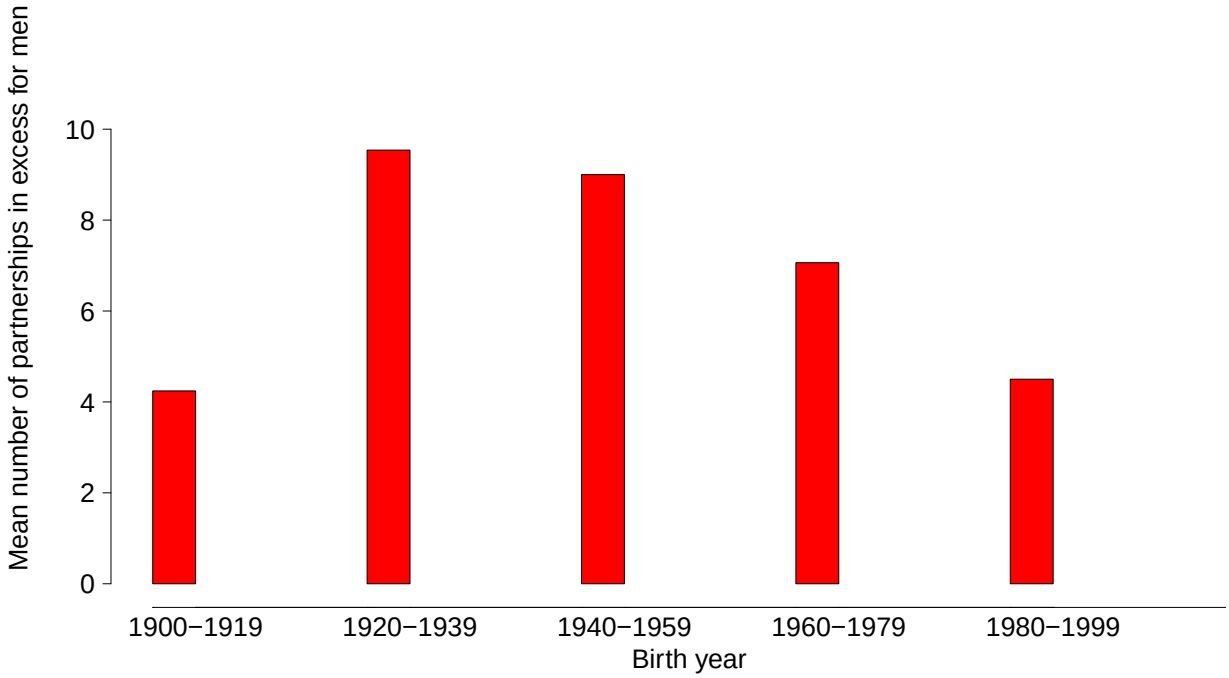


Figure C.30 – **Mean number of partnerships formed by men in excess (in the model population without correction) compared to partnerships formed by women by birth cohorts of the male partner**.

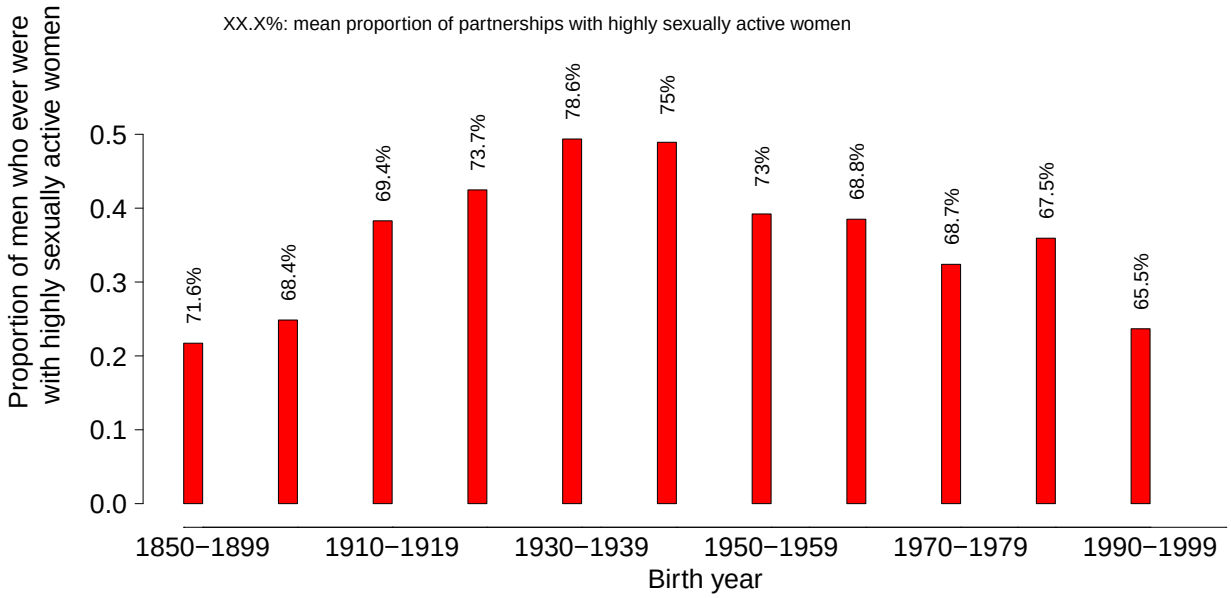


Figure C.31 – **Proportion of men who formed partnerships with high-risk women by birth cohorts in the model population**.

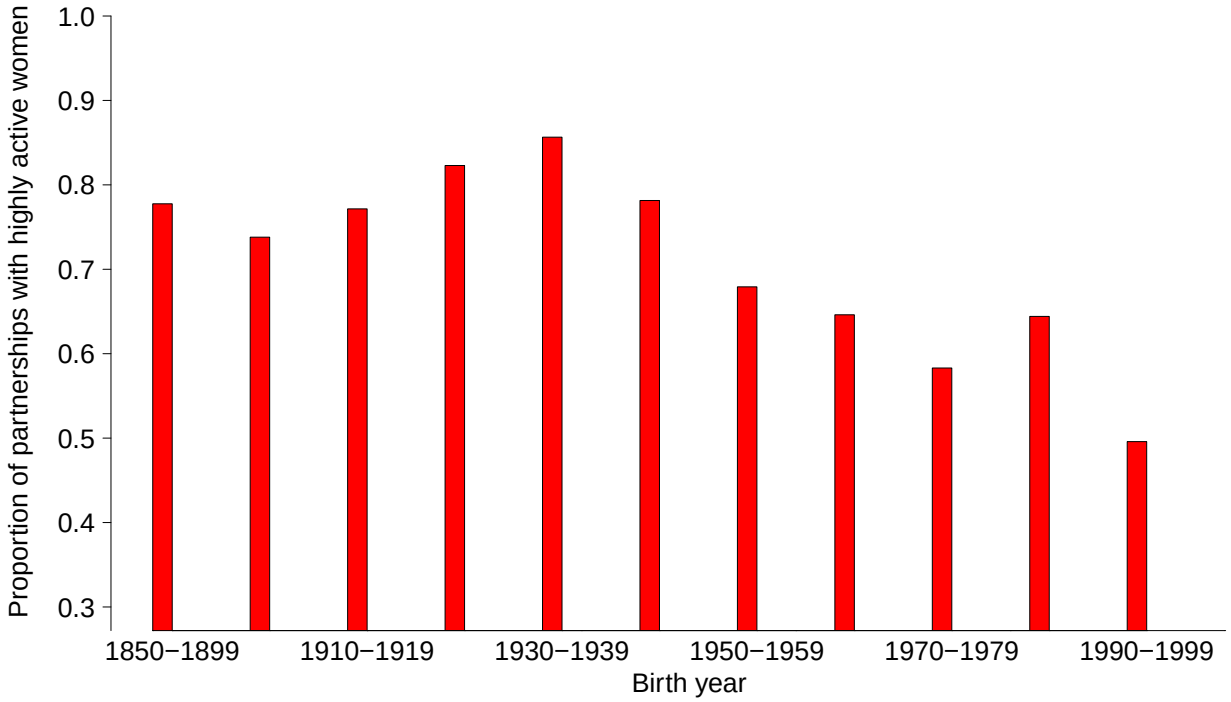


Figure C.32 – **Proportion of all partnerships that were formed with high-risk women by birth cohorts in the model population**.

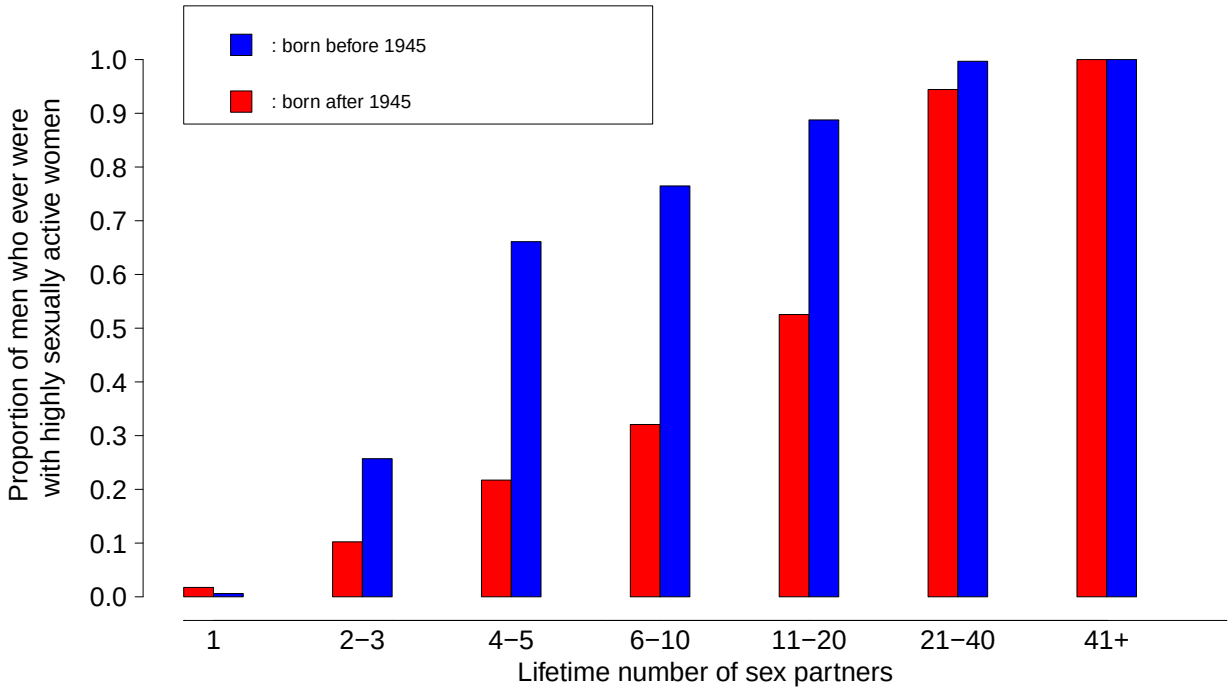


Figure C.33 – **Proportion of men who formed partnerships with high-risk women by lifetime number of sexual partners in the model population**.

C.8.2 Scenario 2: Men overestimate, women underestimate

	Definitions
Variables	
A	age of at interview
LP	lifetime number of sex partners
Transformations	
t_{LP}	piece-wise linear decomposition with knots at 1, 4, 10, and 40 partners
t_A	piece-wise linear decomposition with knots at 20, 30, and 40 years

Table C.15 – **Definitions of variables and transformations used to describe the models to identify men who overestimate and women who underestimate their lifetime number of sex partners**

$\boldsymbol{Models\ RPB1 - RPB3}$

Model $RPB1$ is a logistic model used to determine if a man has overestimated or a woman has underestimated their lifetime number of sex partner. Model $RPB2$ is a negative binomial zero-inflated regression model used to determine the number of additional sex partners a woman has. Model $RPB3$ is a beta regression model used to determine the proportion of partnerships that was overestimated by a man:

$$log(E(response)) = t_{LP}(LP) \times t_A(A)$$

Model fit figures

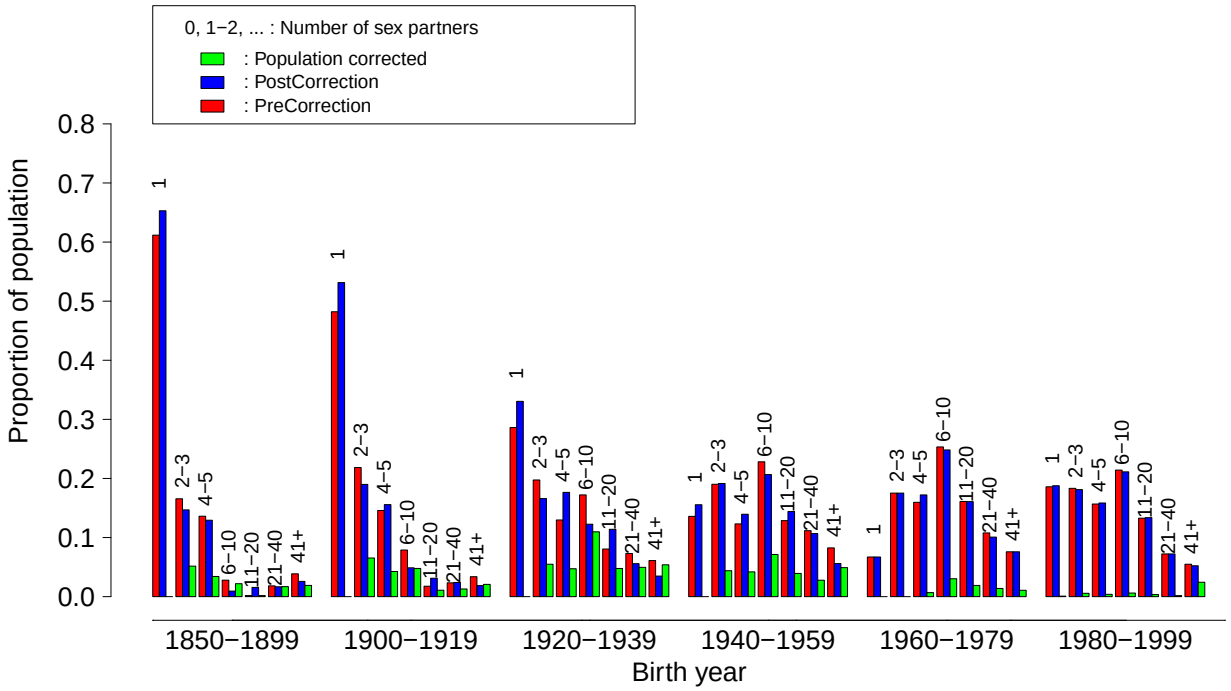


Figure C.34 – **Correction of the overestimation of the lifetime number of sex partners of men by birth cohorts**.

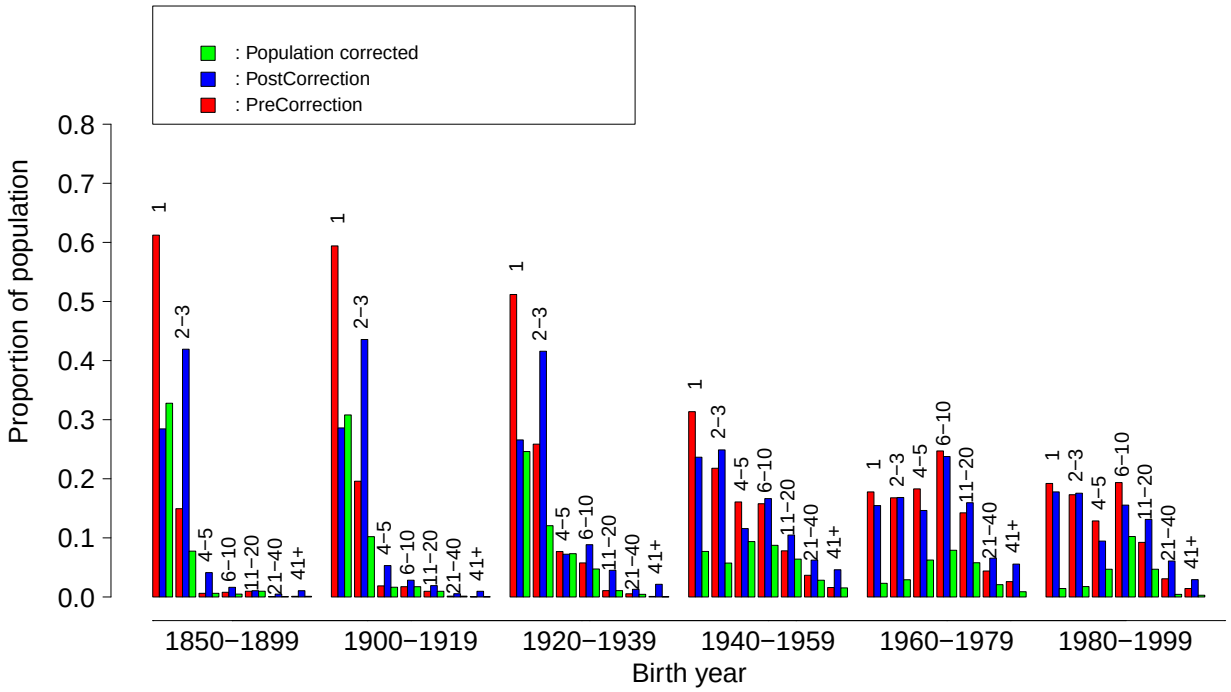


Figure C.35 – **Correction of the underestimation of the lifetime number of sex partners of women by birth cohorts**.

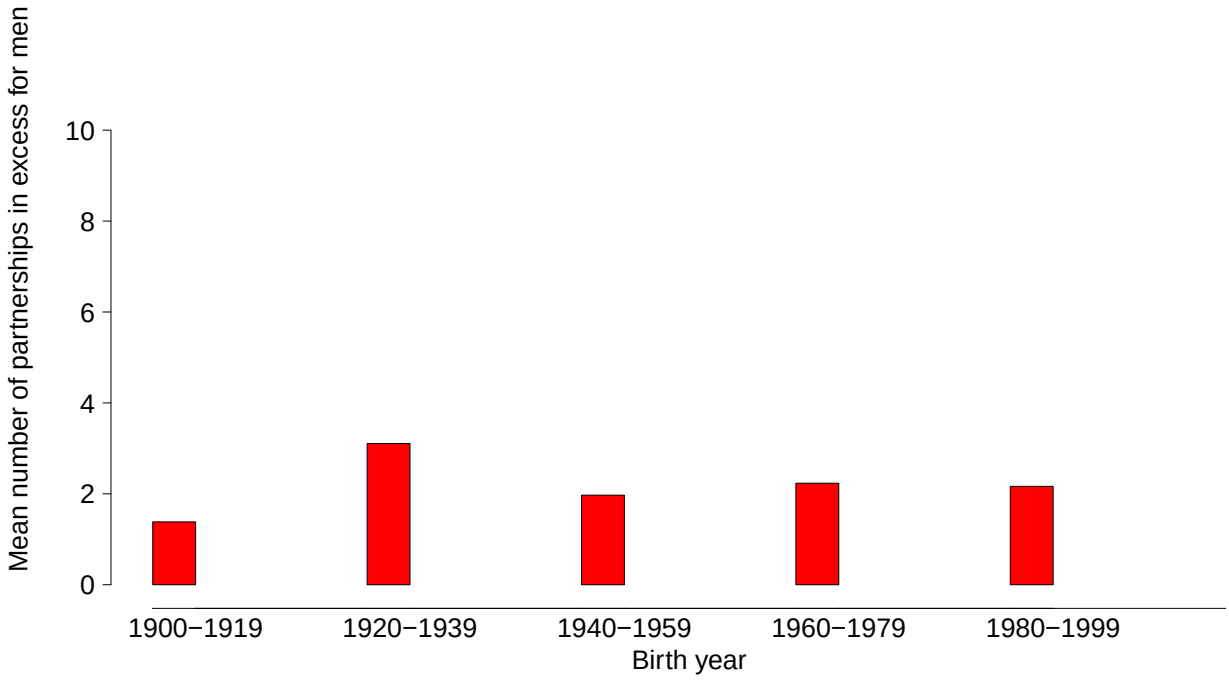


Figure C.36 – **Mean number of partnerships formed by men in excess (in the model population after correction described in scenario 2) compared to partnerships formed by women by birth cohorts of the male partner**.

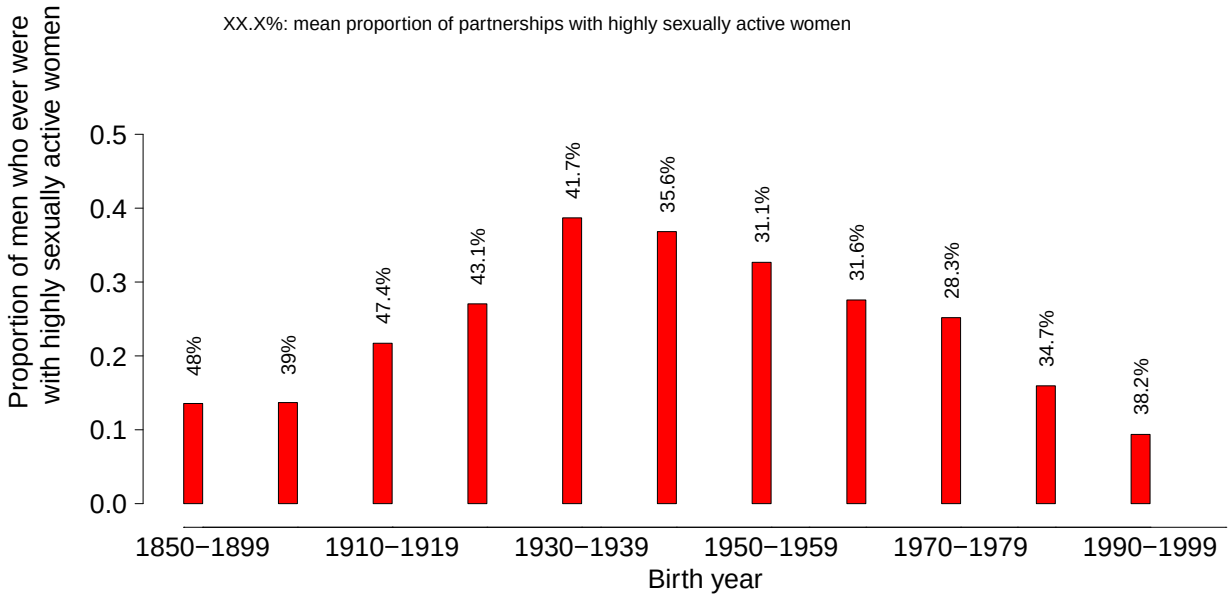


Figure C.37 – **Proportion of men who formed partnerships with high-risk women by birth cohorts in the model population**.

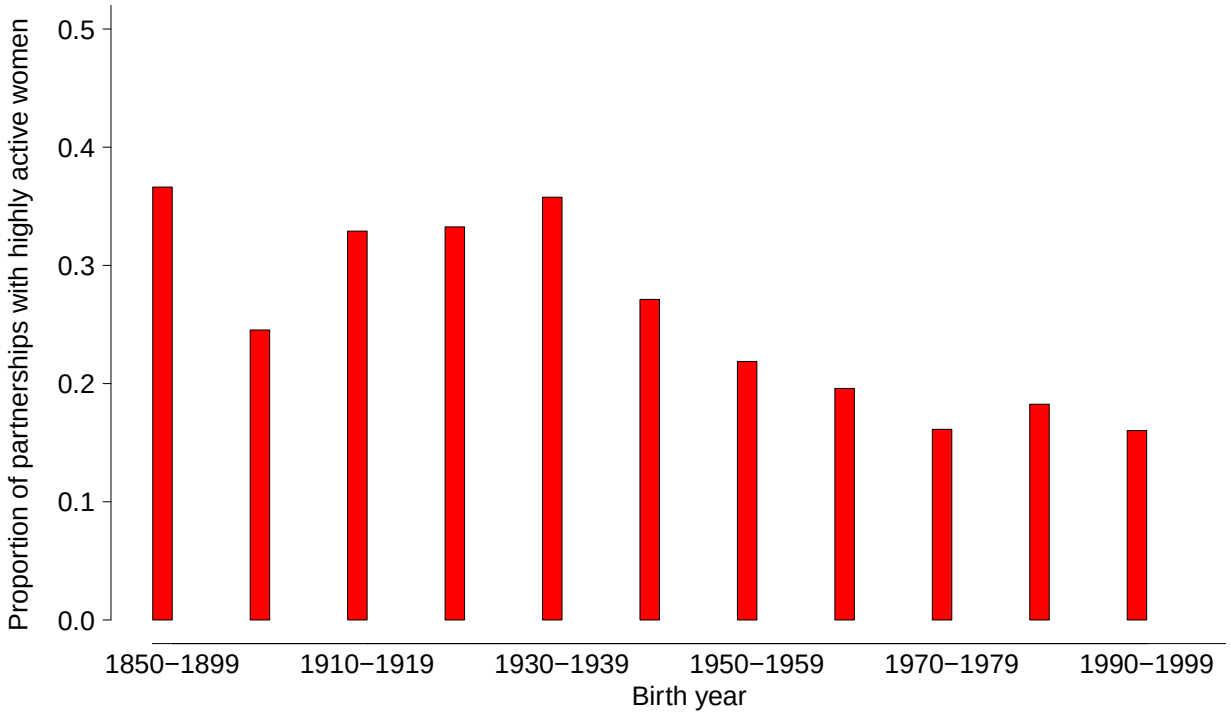


Figure C.38 – **Proportion of all partnerships of men that were formed with high-risk women by birth cohorts in the model population**.

C.9 Who mixes with whom: non-cohabiting partnerships

C.9.1 Level of sexual activity of the partner

Female and male partners outside the model population are given a category of lifetime number of partners ($1-4$, $5-10$, $11-20$ and $21+$). Male partner inside the model population also have their lifetime number of partners category determined before the sampling of the male partner. To determine the category of lifetime number of partners, we used three logistic models for men and three logistic models for women.

	Definitions
Variables	
AD	age difference: age of man minus age of woman
A_s	age of participant at start of partnership
B	year of birth
LP	lifetime number of sex partners
Transformations	
t_B	piece-wise linear decomposition with knots at 1920, 1940, 1960, and 1980 years
t_{A_s}	piece-wise linear decomposition with knots at 20, 25, and 35 years
t_{AD}	categorization as ≤ -6, $]-6,-3]$, $]-3,-1]$, $]-1$,1],]1,3],]3,8], and >8 years
t_{LP}	categorization as $1-4$, $5-10$, $11-20$, and $21+$ partners

Table C.16 – **Definitions of variables and transformations used to describe the models of the lifetime number of partners of the non-cohabiting partners**

$\boldsymbol{Models\ LFNPP1-LFNPP3}$

Models $LFNPP1-LFNPP3$ are logistic models used to determine the category of lifetime number of partners for the non-cohabiting partner: $t_{LP}(LP)$. The models follow the same conditional structure as in C.5.1:

$$logit(E(response)) = t_{A_s}(A_s) \times t_{AD}(AD) \times t_B(B)$$

C.10 Individual-based dynamic model of STI and subsequent development of cancer

Parameters	Scenarios		
	High-risk group	*Report bias men and women*	*Report bias men only*
Probabilities of transmission per partnership			
Sexual intercourse: Genital → Genital	0.78	0.78	0.78
Sexual intercourse: Genital → Oral	0	0	0
Oral Sex: Genital → Oral	0	0	0
Natural history of infection			
Men and Women: Clearance rate		1/18 months^{-1}	
Men: Probability of developing systemic natural immunity		10%	
Women: Probability of developing systemic natural immunity		35%	
HPV-related cancers			
Mean duration from infection to cervical cancer (years)	25	25	25
Standard deviation of duration from infection to cervical cancer	9	9	9
Probability for an infection to transition to cervical cancer	0.017	0.009	0.019
Mean duration from infection to oropharyngeal cancer (years)	41	40	41
Standard deviation of duration from infection to oropharyngeal cancer	9	9	9
Probability for an infection to transition to oropharyngeal cancer	0.0017	0.0014	0.0014

Table C.17 – **Median values of parameters of HPV transmission for scenario uni-site**

Parameters	Scenarios		
	High-risk group	*Report bias men and women*	*Report bias men only*
Probabilities of transmission per partnership			
Sexual intercourse: Genital → Genital	0.78	0.78	0.78
Sexual intercourse: Genital → Oral (with oral sex)	0	0	0
Oral Sex: Genital → Oral	0.175	0.400	0.220
Natural history of infection			
Men and Women: Clearance rate	1/18 months^{-1}		
Men: Probability of developing systemic natural immunity	10%		
Women: Probability of developing systemic natural immunity	35%		
HPV-related cancers			
Mean duration from infection to cervical cancer (years)	25	25	25
Standard deviation of duration from infection to cervical cancer	9	9	9
Probability for an infection to transition to cervical cancer	0.017	0.009	0.019
Mean duration from infection to oropharyngeal cancer (years)	41	41	41
Standard deviation of duration from infection to oropharyngeal cancer	9	9	9
Probability for an infection to transition to oropharyngeal cancer	0.134	0.075	0.099

Table C.18 – **Median values of parameters of HPV transmission for scenario multi-site with oral sex**

Parameters	Scenarios		
	High-risk group	*Report bias men and women*	*Report bias men only*
Probabilities of transmission per partnership			
Sexual intercourse: Genital → Genital	0.78	0.78	0.78
Sexual intercourse: Genital → Oral	0.084	0.144	0.102
Oral Sex: Genital → Oral	0	0	0
Natural history of infection			
Men and Women: Clearance rate	1/18 months^{-1}		
Men: Probability of developing systemic natural immunity	10%		
Women: Probability of developing systemic natural immunity	35%		
HPV-related cancers			
Mean duration from infection to cervical cancer (years)	25	25	25
Standard deviation of duration from infection to cervical cancer	9	9	9
Probability for an infection to transition to cervical cancer	0.017	0.009	0.019
Mean duration from infection to oropharyngeal cancer (years)	40	40	40
Standard deviation of duration from infection to oropharyngeal cancer	9	10	9
Probability for an infection to transition to oropharyngeal cancer	0.099	0.004	0.097

Table C.19 – **Median values of parameters of HPV transmission for scenario multi-site without oral sex**

C.11 The practice of oral sex

$Model\ OSP1$

Model $OSP1$ is used to determine whether an individual practices oral sex:

$$logit(E(response)) = t_B(B)$$

with $t_B(B)$ being a piece-wise linear decomposition of year of birth with a knot at 1940.

$Model\ OSP2$

Model $OSP2$ is used to determine whether an individual who practices oral sex (as determined in Model $OSP1$) will perform oral sex with a given partner:

$$logit(E(response)) = P \times (\beta_1 \times C + \beta_0)$$

with P being an indicator variable of whether the partnership is with a woman from the high-risk group, C being indicator variable of whether the partnership is a cohabitation.

C.11.1 Model fit figures

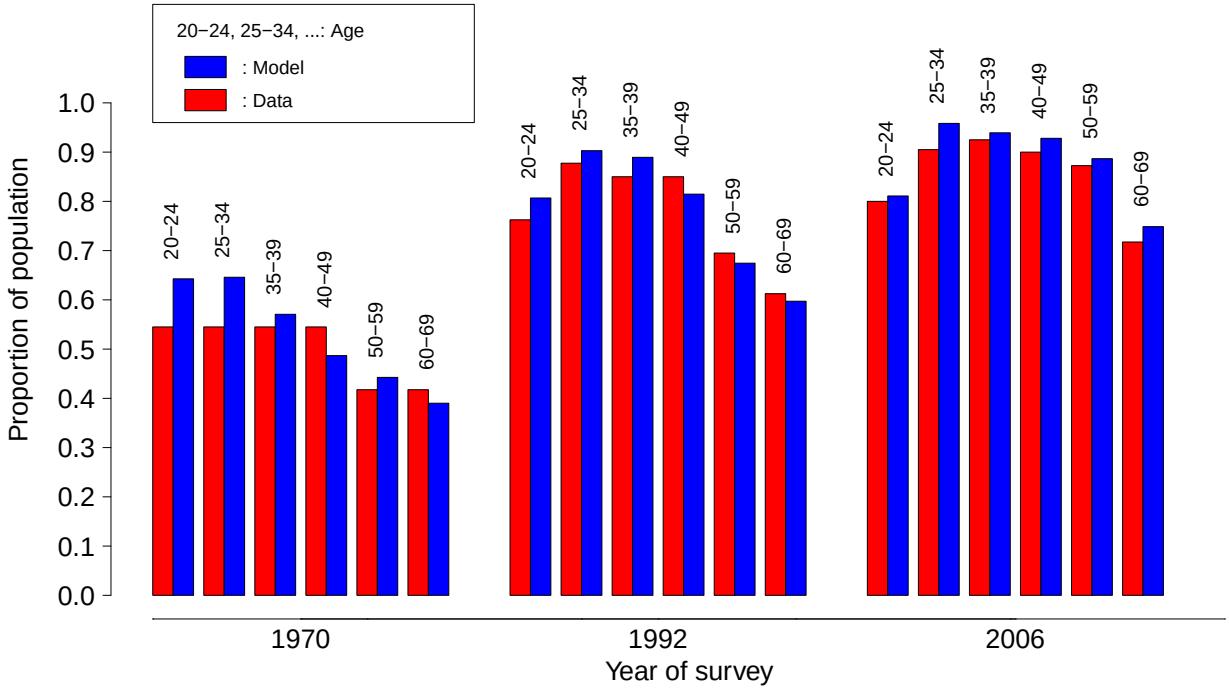


Figure C.39 – **Proportion of women who have ever had cunnilingus performed on them by year of birth: model simulations of Scenario 1 vs data from three surveys from France**.

	Data	Scenario 1	Scenario 2	Scenario 3
Proportion of men who performed cunnilingus (%)	90.9	89.5	90.1	89.3

Table C.20 – **Proportion of men aged 40-44 years old between 2011-2015 who have performed cunnilingus on a woman: model simulations vs data from three France surveys**

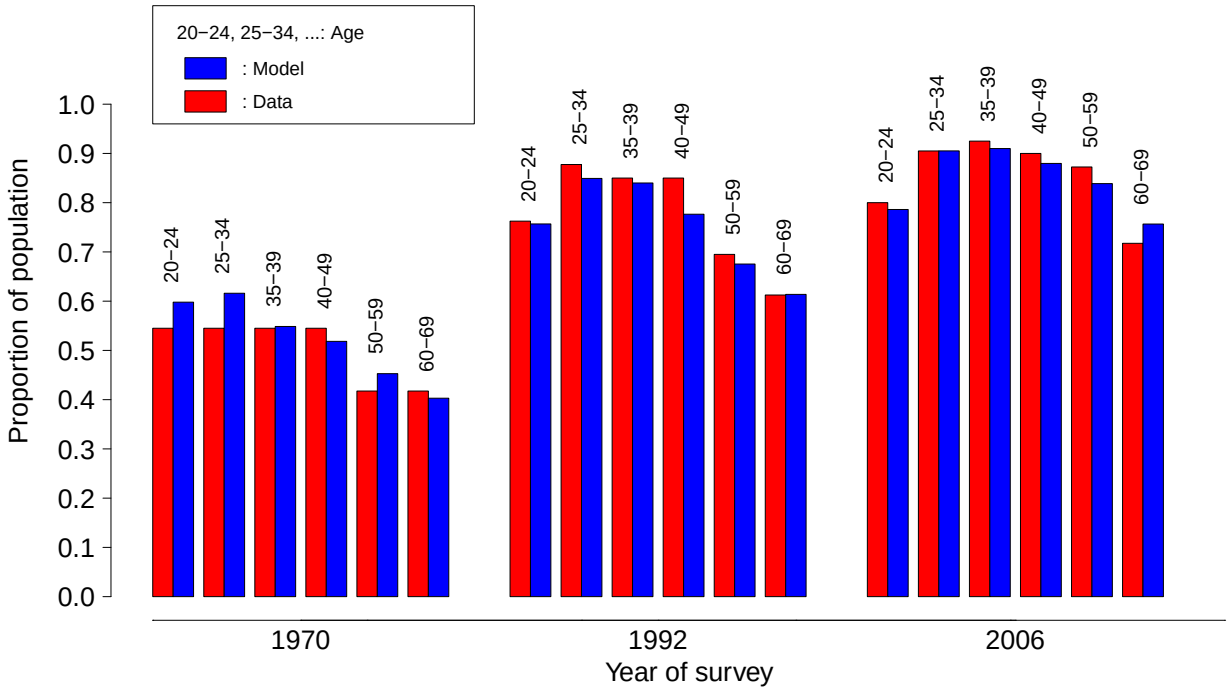


Figure C.40 – **Proportion of women who have ever had cunnilingus performed on them by year of birth: model simulations of Scenario 2 vs data from three surveys from France**.

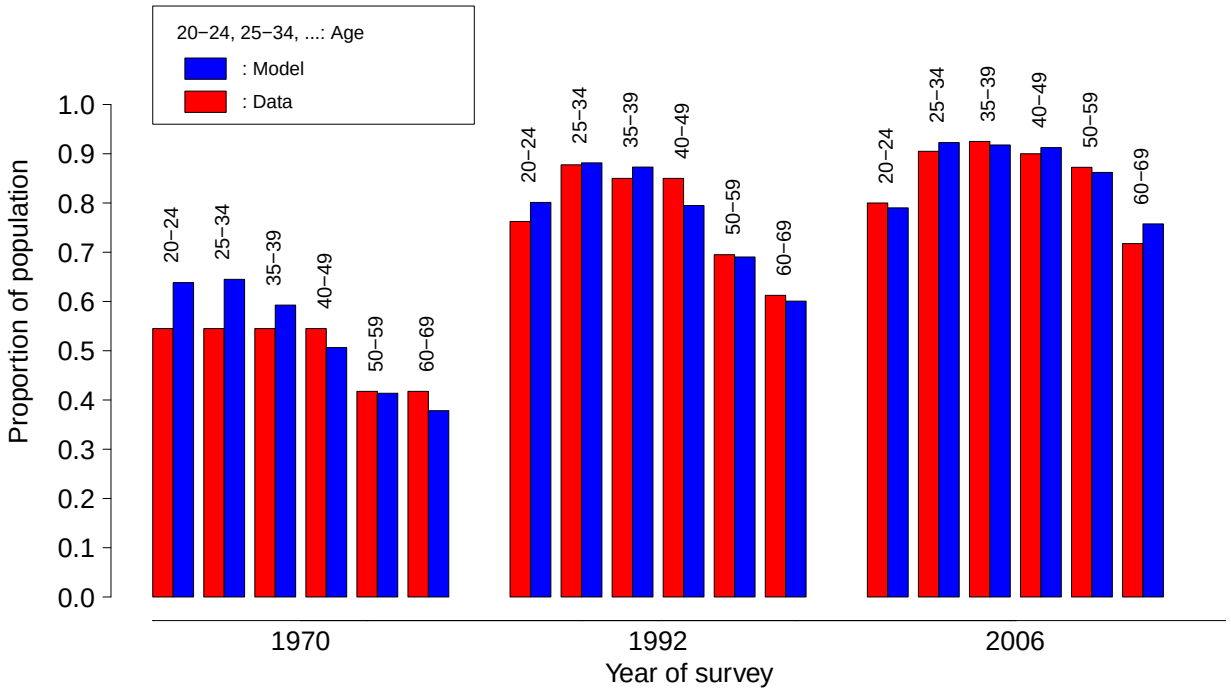
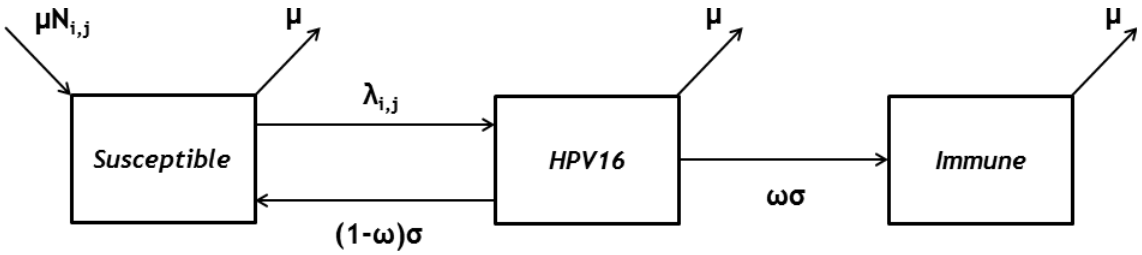


Figure C.41 – **Proportion of women who have ever had cunnilingus performed on them by year of birth: model simulations of Scenario 3 vs data from three surveys from France**.

Appendix D

Supplementary materials. Modelling multi-site transmission of the human papillomavirus and its impact on vaccination effectiveness

D.1 Model flow chart



μ: Exit rate (mortality and aging)

$N_{i,j}$: Total population in class *i,j* (*i=yes,no* for the smoking status and *j=high,low* for the level of sexual activity

ω: Probability of acquiring natural immunity

σ: Clearance rate

$\lambda_{i,j}$: Force of infection in class *i,j*

Figure D.1

Table D.1 – **Base-Case Model Parameter Values**

Population parameters	**Symbol**	**Base-case**	**Reference**
Natural mortality rate (per year)	μ	1/30	Assumption
Probability of transmission per partnership	π	75%	[186]
Average duration of infection (years)	$D = \frac{1}{\sigma}$	1.2	[186]
Probability of developing natural immunity after clearance of infection	ω	30%	[186]
Proportion of smokers in the population		20%	[249]
Proportion of individuals in the high activity class	α_S	1.2%	[249]
Risk ratio of being highly sexually active if smoker compared to non-smoker	α_{NS}	7.5%	[249]
Rate of new sexual partners per year in sexual activity class i	n_i	$n_1 = 1.2$ $n_2 = 7.0$	[249]
Sexual assortativity parameter	ϵ_1	0.4	Assumption
Smoking assortativity parameter	ϵ_2	0.8	Assumption
Parameter of correlation between the two types of assortativity	c	1	Assumption

D.2 Parameter values

D.3 Model equations

$$\frac{\partial Y_{i,j}}{\partial t} = -\sigma Y_{i,j} - \mu Y_{i,j} + n_i \pi X_{i,j} \left(p_{i,j}(1,1)\frac{Y_{1,1}}{N_{1,1}} + p_{i,j}(1,2)\frac{Y_{1,2}}{N_{1,2}} + p_{i,j}(2,1)\frac{Y_{2,1}}{N_{2,1}} + p_{i,j}(2,2)\frac{Y_{2,2}}{N_{2,2}} \right) \tag{D.1}$$

$$\frac{\partial Z_{i,j}}{\partial t} = \omega\sigma Y_{i,j} - \mu Z_{i,j} \tag{D.2}$$

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), n_i= rate of new sexual partner acquisition, $X_{i,j}$= susceptible, $Y_{i,j}$= infected, $Z_{i,j}$= immune, $N_{i,j}$= total population ($X_{i,j}$+$Y_{i,j}$+$Z_{i,j}$), σ=clearance rate, μ=mortality rate, π=probability of transmission per partnership, ω=probability of developing natural immunity after clearance, $p_{i,k}(j,l)$=proportion of partnerships of individuals in class (i,k) that are with individuals in class (j,l).

D.4 Mixing equations

The quantities $p_{i,k}(j,l)$ 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:

$$p_{i,k}(j,l) = a\delta_{i,j}\delta_{k,l} + b\delta_{i,j}\frac{N_{j,l}n_l}{\sum_p N_{i,p}n_p} + c\delta_{k,l}\frac{N_{j,l}n_l}{\sum_m N_{m,k}n_k} + (1-a-b-c)\frac{N_{j,l}n_l}{\sum_m\sum_p N_{m,p}n_p} \quad \text{(D.3)}$$

Where a,b,c are non-negative numbers such that $a+b+c \leq 1$ and is equal 1 if $i = j$ and 0 otherwise. Other notation is as previously. These mixing equations are algebraically always balanced for any non-negative a,b,c such that $a + b + c \leq 1$ because they are a convex sum of solutions (each term of the sum can be seen to be a solution).

Formula D.4 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 D.4 as probabilities of event, we can define the four probabilities as obtained by two probabilities: the probability of selecting assortatively for smoking status, ϵ_1, and for sexual activity, ϵ_2. The alternative to selecting assortatively is selecting randomly. Thus, the probability a becomes $\epsilon_1\epsilon_2$. We can further introduce a correlation between the two events using a parameter, α (α is not the correlation however). The new formula would then be:

$$\begin{aligned} p_{i,k}(j,l) &= \epsilon_1\alpha\epsilon_2\delta_{i,j}\delta_{k,l} + (1-\epsilon_1)\epsilon_2\left(\tfrac{1-\alpha\epsilon_1}{1-\epsilon_1}\right)\delta_{i,j}\frac{N_{j,l}n_l}{\sum_p N_{i,p}n_p} + \epsilon_1(1-\alpha\epsilon_2)\delta_{k,l}\frac{N_{j,l}n_l}{\sum_m N_{m,k}n_k} \\ &+(1-\epsilon_1)\left(1-\epsilon_2\left(\tfrac{1-\alpha\epsilon_1}{1-\epsilon_1}\right)\right)\frac{N_{j,l}n_l}{\sum_m\sum_p N_{m,p}n_p} \end{aligned} \quad \text{(D.4)}$$