Réseaux de contacts sexuels et infections par le VIH : apport de leur modélisation et de la simulation dans la compréhension de l'épidémie

Olivier Robineau

▶ **To cite this version:**

Olivier Robineau. Réseaux de contacts sexuels et infections par le VIH : apport de leur modélisation et de la simulation dans la compréhension de l'épidémie. Médecine humaine et pathologie. Sorbonne Université, 2020. Français. ⟨NNT : 2020SORUS466⟩. ⟨tel-03994607⟩

HAL Id: tel-03994607

https://theses.hal.science/tel-03994607v1

Submitted on 17 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

THESE DE DOCTORAT DE
SORBONNE UNIVERSITE

ECOLE DOCTORALE PIERRE LOUIS DE SANTE PUBLIQUE A PARIS : EPIDEMIOLOGIE ET SCIENCES DE L'INFORMATION BIOMEDICALE

ED393

Spécialité

Epidémiologie et biostatistique

Présentée par

M. Olivier ROBINEAU

Pour obtenir le grade de

DOCTEUR de SORBONNE UNIVERSITE

Sujet de la thèse :

RESEAUX DE CONTACTS SEXUELS ET INFECTIONS PAR LE VIH: APPORT DE LEUR MODELISATION ET DE LA SIMULATION DANS LA COMPREHENSION DE L'EPIDEMIE

soutenue le 17 Juilllet 2020

devant le jury composé de :

M. Pierre-Yves Boelle Directeur de thèse

M. Bruno Spire Rapporteur

M. Viet-Chi Tran Rapporteur

Mme Karine Lacombe Examinatrice

Mme Annie Velter Examinatrice

M. Yazdan Yazdanpanah Examinateur

Sorbonne Université
Bureau d'accueil, inscription des doctorants et base de données
Esc G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à MON : 01 42 34 68 41
Tél. pour les étudiants de MOO à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

Laboratoire de rattachement

Sorbonne Université, INSERM
Institut Pierre Louis d'Épidémiologie et de Santé Publique
Directreur : Monsieur Fabrice Carrat

Équipe 1 – Surveillance et modélisation des maladies transmissibles (SUMO)
Responsable : M. Pierre-Yves Boëlle
Faculté de médecine Pierre et Marie Curie, site Saint-Antoine 27, rue de Chaligny
75571 Paris cedex 12
France

Remerciements

Je souhaiterais exprimer ma profonde reconnaissance à Pierre-Yves Boëlle pour tout ce qu'il m'a appris au fil des années, du M2 à la thèse et pas uniquement sur les réseaux de contact. De l'élaboration de projets à l'écriture du travail en passant par la recherche de subventions, la formation a été complète.

Je remercie aussi tous les membres de l'unité INSERM 1136 et plus particulièrement Gilles pour les conseils, les cafés et la musique, les thésard Thomas, Julien et Tristan pour leur soutien, le fourmillement d'idées et les boucheries.

J'exprime toute ma gratitude envers Annie Velter et l'équipe de Santé Publique France qui m'a permis l'accès aux données des différentes enquêtes PREVAGAY.

Je remercie l'équipe brésilienne de Fiocruz à Rio de Janeiro, de l'université de Sao Paolo et de Fortaleza pour l'accès aux bases de données nationales et régionales et également pour leur accueil sur place : André Perissé, Ines Dourado, Ligia Kerr.

L'ensemble des membres de l'équipe du Service Universitaire des Maladies Infectieuses et du Voyageur de Tourcoing a été un profond moteur à la rédaction de cette thèse et je les remercie pour leur soutien. Merci Yazdan (je te mets dans l'équipe tourquinoise) de m'avoir mis le pied à l'étrier, d'être toujours resté disponible et de m'avoir orienté vers Pierre-Yves. Merci Eric pour ton indéfectible soutien et pour l'apprentissage permanent à tes côtés.

Je remercie mes parents et mon frère pour m'avoir supporté tout le long de mes interminables études (ce n'est peut-être pas fini…)

Marie, Raphaël, Louis, votre amour a été le moteur de ce travail, merci pour tout. Infiniment.

Plan de thèse

Tables et figures 5
Résumé 6
Introduction 9
1. Réseaux, épidémie et VIH : approche analytique 14
1.1 Les réseaux 14
1.1.1 Définitions 14
1.1.2. Les différents types de données de réseaux 15
1.2. Utilisation des données de réseaux en épidémiologie analytique 18
1.2.1. Généralités 18
1.2.2. Évaluation du risque d'acquisition du VIH et variables de réseaux 19
2. Impact des réseaux dans les processus épidémiques et leur prévention : modélisation de réseaux et simulation d'épidémie 28
2.1. Généralités 28
2.2. Les outils d'élaboration de la simulation d'épidémie au sein de réseaux 31
2.2.1 Généralités 31
2.2.2. Élaboration de graphes 32
2.2.3. Les ERGMs : Un outil mathématique de modélisation de réseau 33
2.2.4. L'ERGM et sa composante dynamique (105,110,111) 34
2.2.5. Les modèles de transmission du VIH 35
2.3 Modélisation des réseaux de contacts sexuels afin d'expliquer l'évolution de l'épidémie liée au VIH et impact des mesures de prévention 37
2.3.1 Généralités 37
2.3.2. Modélisation des réseaux de contacts sexuels dans une population HSH française, impact de sa structure sur l'efficacité des mesures de prévention 38
2.3.2.1 introduction 38
2.3.2.2. Étude PREVAGAY 2009 et données de réseaux (119) 40
2.3.2.3 Le modèle de transmission 41
A Probabilité d'infection sans intervention 42
B Intervention 42
C Définition des transmetteurs 44
D Évolution de la charge viral (CV) 44
E Évolution des CD4 46
F Probabilité d'être sous traitement 47
G. Évènements classant SIDA sans traitement (table 4) 47
H Réalisation des tests de dépistage 47
2.3.2.4 Article publié 49

2.3.2.5 Résultats principaux 64
2.3.2.6. Critique du modèle de l'article 65
3. Utilisation des réseaux dans l'estimation d'une prévalence de maladies 67
3.1. Généralités sur les méthodes d'échantillonnage par les pairs 67
3.2 Respondent Driven Sampling (RDS) 68
3.3 Outils de redressement des données de type RDS 71
3.3.1. Outils historiques 71
3.3.2. Méthodes d'estimation utilisant les réseaux 73
3.3 Article 78
3.5 Principaux résultats et réflexion sur la qualité des données 90
4. Synthèse 92
5. Ouverture : les réseaux de contacts comme outils clefs de l'action 95
Conclusion 97
Bibliographie 98

Tables et figures

Figure 1 Définition générique d'un réseau et de ses constituants et application à l'épidémiologie. 15
Figure 2 Représentation d'un réseau sociocentrique (A) ; ego centrique (B) et mixte (C). 16
Figure 3 Variables de réseaux ayant un impact dans la transmission. 23
Figure 4 Représentation d'un réseau représentatif d'un individu infecté (cas) et d'un individu non infecté (témoins). 25
Figure 5 Déroulement d'une étude de type RDS. 69
Figure 6 Algorithme permettant l'estimation d'une prévalence selon Gile (en orange). Méthode ajoutée permettant l'estimation d'une homophilie de travail lorsque ce paramètre n'est pas connu (en jaune). 75
Figure 7 Comparaison des différents estimateurs 76
Figure 8 Stratégies de recrutement dans des programmes de prévention. 96

Tableau 1 Résultat de la régression comparant les données de réseaux des individus infectés aux individus non infectés. 24
Tableau 2 Variables conditionnant la probabilité de transmission 43
Tableau 3 Variables définissant la valeur de la charge virale de manière mensuelle. 45
Tableau 4 Variables définissant la probabilité d'évènement classant SIDA 48
Tableau 5 Prérequis et hypothèse d'échantillonnage théoriquement nécessaires à l'estimation d'un paramètre à partir de RDS. 73

Résumé

La notion de réseaux de contacts est majeure en épidémiologie des maladies transmissibles. Les données de réseaux peuvent être utilisées 1) à des fins analytiques, dans l'objectif d'intégrer des variables liées au réseau dans le risque d'acquérir une pathologie, 2) à des fins de simulation, afin d'évaluer l'importance des variables de réseaux dans un processus épidémique et 3) à des fins de redressement de données épidémiologiques issues d'échantillons non représentatifs échantillonnés au sein d'un réseau. L'étude des réseaux et leur impact dans l'évolution d'une épidémie a pris son essor avec l'apparition du VIH. Ce virus se transmet essentiellement par voie sexuelle en suivant les liens formant un réseau de contacts dit sexuels. Après avoir dressé le panorama bibliographique sur l'utilisation des variables de réseaux dans l'analyse du risque de la transmission du VIH nous développons deux travaux utilisant la modélisation et la simulation de réseaux de contacts sexuels dans des populations d'hommes ayant des rapports sexuels avec des hommes (HSH).

Le premier travail détaille l'usage des réseaux dans la simulation de l'épidémie liée au VIH et leur impact sur l'effet des mesures de prévention. A partir des données de l'étude PREVAGAY portant sur les hommes ayant des rapports sexuels avec des hommes visitant des lieux de rencontre parisiens, des réseaux de contacts sexuels réalistes ont été développés. Des processus épidémiques ont été générés dans ces réseaux et comparés à des réseaux théoriques. Nous démontrons que la structure du réseau influe non seulement sur l'évolution de la courbe épidémique mais également sur l'efficacité des mesures de prévention individuelles telles que la prophylaxie de préexposition. Son efficacité peut être doublée selon les caractéristiques réseaux des individus entrant dans ce type de programme de prévention. Ce travail confirme la nécessité de colliger des données de réseau en France en vue de préciser le modèle développé.

Le deuxième travail généralise une méthode d'estimation de la prévalence du VIH réseaux à partir de données d'observation partielle et utilisant la modélisation de réseaux. Cet outil était antérieurement peu applicable aux données réelles. Il démontre la robustesse de la méthode face aux biais classiquement retrouvés dans les données récoltées par échantillonnage orientés par les pairs (respondent driven sampling (RDS). L'application de cette méthode d'estimation a été comparée aux méthodes historiques d'estimation à partir de données réelles issues d'une étude de type RDS réalisé chez les HSH vivant dans les grandes zones urbaines au Brésil. Le travail confirme l'inquiétante prévalence de l'infection par le VIH dans cette population malgré de grandes disparités territoriales. Il ouvre vers d'éventuelles comparaisons de données historiques en vue de comprendre l'évolution de l'épidémie liée au VIH au Brésil.

L’ensemble de ce travail suggère que les caractéristiques des réseaux peuvent être utilisées à des fins d’interventions dans le but de mettre en œuvre des politiques de préventions rapidement efficientes.

Abstract

Contact networks is a major concept in the epidemiology of communicable diseases. Network data can be used 1) for analytical purposes, with the aim of integrating network variables into the risk of acquiring a disease, 2) for simulation purposes, to assess the importance of network variables in the epidemic process, and 3) for the adjustment of epidemiological data from non-representative samples sampled within a network.
The study of networks and their impact in the evolution of an epidemic took off with the emergence of HIV. This virus is essentially transmitted by sexual transmission by following the links forming a so-called “sexual contact network”.

After a bibliographical overview of the use of network variables in the analysis of the risk of HIV transmission, we develop two research work using modelling and simulation of sexual contact networks in populations of men who have sex with men (MSM).
The first work details the use of networks in the simulation of the HIV epidemic and their impact on the effect of prevention measures. Based on data from the PREVAGAY study of men who have sex with men visiting Parisian venues, realistic networks of sexual contacts were developed. Epidemic processes were generated in these networks and compared to theoretical networks. We show that the structure of the network influences not only the evolution of the epidemic curve but also the effectiveness of individual prevention measures such as pre-exposure prophylaxis. Its effectiveness can be doubled depending on the network characteristics of the individuals entering this type of prevention program. This work confirms the need to collect network data in France in order to refine the model developed.
The second work generalizes a method for estimating HIV prevalence using network modelling based on partial observational data that was previously not very applicable to real data. It demonstrates the robustness of the method in the face of the biases classically found in data collected by respondent driven sampling (RDS). The application of this estimation method was compared to historical methods of estimation from real data from an RDS-type study conducted among MSM living in large urban areas in Brazil. The work confirms the worrying prevalence of HIV infection in this population despite large territorial disparities. It opens the door to possible comparisons of historical data in order to understand the evolution of the HIV epidemic in Brazil.
All these results suggest that the characteristics of the networks can be used for intervention purposes with the aim of implementing prevention policies in an efficient manner.

Introduction

Les maladies infectieuses peuvent être transmises d'un individu à un autre de différentes manières, par voie aérienne, orale, sexuelle, sanguine ou par contact, direct ou indirect (1). Quel que soit le mode de transmission, le fait qu'un individu transmette un agent pathogène à un autre crée une chaine de transmission. Cette chaine se développe au sein d'un réseau de contact. Ce réseau est notamment défini par la nature des liens entre les individus qui permet la transmission de l'agent pathogène. Ainsi, l'hypothèse selon laquelle la structure des réseaux de contacts peut avoir un impact dans la diffusion d'une infection a été développée (2). L'étude des réseaux de contact ou « réseaux sociaux » est en plein essor depuis la fin des années 1980 de par l'intérêt grandissant des épidémiologistes pour des outils mathématiques issus de la sociologie quantitative, le développement des capacités computationnelles permettant leur application et l'apparition d'une pathologie transmissible, le VIH (3,4). L'étude des réseaux sociaux et leur application aux phénomènes épidémiques comportent ainsi plusieurs challenges.

Tout d'abord, l'individualisation du réseau, qui semble avoir un impact dans la transmission d'un agent pathogène, est difficile. Il convient de réussir à le définir. Pour exemple, un réseau formé par des liens amicaux peut être en partie commun à un réseau de contacts professionnels, et à un réseau de contacts sexuels. Les individus peuvent ainsi avoir plusieurs liens différents entre eux. Individualiser un type de lien pour étudier la transmission d'une maladie est donc réducteur mais indispensable dans la réalisation d'une étude. Il existe donc une grande complexité dans la détermination du réseau à étudier et des simplifications sont à déterminer d'emblée (5). Dans l'étude de la transmission du VIH et des autres infections sexuellement transmissibles (IST), la notion de « réseau de contacts sexuels » a été élaborée, défini par les individus ayant eu des rapports sexuels entre eux.

Ensuite, la collecte des données de réseaux est longue et souvent fastidieuse voire décevante. Comme cela est décrit plus en détail dans le premier chapitre de cette thèse, la collecte des données permettant de décrire les réseaux nécessite l'élaboration d'enquêtes difficiles à mettre en œuvre, car elles ne touchent pas que les individus inclus mais également leurs contacts. Les enquêtes les plus simples, dites *ego-centrique*s, s'intéressent par exemple à documenter les contacts de l'individu observé (désigné par *ego*) et à décrire ses contacts (désignés comme *alters*). Les enquêtes *socio*-centriques, plus détaillées et plus rares, visent quant à elles à documenter l'ensemble des contacts survenant dans une population, chaque individu étant considéré comme un *ego* et un *alter* potentiel (5).

L'analyse de ces données relève de méthodologies propres et différentes selon l'objectif de la recherche. L'étude de la structure d'un réseau nécessite une description de variables spécifiques développées dans la littérature de la statistique spatiale et de la théorie des graphes. Déterminer ce qui constitue le réseau de manière significative est possible grâce à des outils de régression prenant en compte comme paramètre les variables de réseaux. L'intégration des données issues de réseaux de contacts dans l'évaluation du risque pour des individus de présenter un état ou un autre, nécessite la prise en compte des caractéristiques individuelles de *l'ego*, des caractéristiques des contacts (*alters*), des différences ou des similitudes entre les caractéristiques des *egos* et des *alters* (homophilie) et enfin, des variables caractérisant les structures du réseau. La prise en compte des liens de dépendance entre les variables de l'*ego* et celles de ses *alters*, voire même des variables propres aux *alters* entre eux, rend difficile l'usage d'outils de régression.

Enfin, étant donné que l'étude des réseaux est complexe, tant dans la collecte des données que dans leur analyse, elle ne sera pleinement justifiée que si elle apporte des éléments nouveaux, voire déterminants dans la compréhension d'un processus épidémique. Pour que ces données soient d'intérêt, il faut supposer que la maladie soit transmissible et que des facteurs influençant

son acquisition soient influencés par la structure du réseau. Comme démontré dans des modèles s'intéressant à la transmission de la tuberculose, cette hypothèse sera d'autant plus raisonnable dans les situations où (6–8):

- seuls des contacts proches et spécifiques sont nécessaires à la transmission,
- il existe une hétérogénéité de transmission entre groupe
- la transmissibilité de l'agent pathogène est faible

Les données de réseaux peuvent ainsi être utilisées de différentes manières en épidémiologie. Tout d'abord, des variables individuelles décrivant certains aspects du réseau local perçu par un individu peuvent être intégrées dans la détermination du risque d'acquisition d'une maladie. Ceci a été développé de manière extensive dans l'étude des facteurs de risque d'acquisition du VIH, notamment par l'évaluation de l'impact des partenaires simultanés (« concurrency »)(9). Ensuite, elles sont utilisées à des fin de modélisation de processus épidémique pour développer des réseaux dans lesquels sont simulés des épidémies en vue d'objectiver l'impact de la structure des réseaux dans la transmission et l'extension de l'épidémie (10). Enfin, et plus récemment, les outils d'analyse des réseaux de contacts ont été utilisés pour redresser des estimations de prévalence de maladies, réalisées à partir de méthode d'échantillonnage non aléatoire, influencées par la structure des réseaux (11).

Comme dit plus haut, l'étude des réseaux de contacts et leur impact sur la transmission des maladies a pris un essor particulier dans le cas du VIH, et plus généralement des maladies sexuellement transmissibles (3). L'étude des caractéristiques des réseaux de transmission est ainsi devenue majeure pour comprendre l'hétérogénéité de l'épidémie, au plan mondial comme au plan national entre sous populations (12). Rappelons qu'en 2018, 1,7 millions [1,4 millions - 2,3 millions] de personnes ont été infectées par le VIH dans le monde (13). En France, 6200

personnes ont été diagnostiquées infectées par le VIH en 2018. Le nombre d'infections tend à diminuer depuis 2018 en France. Cependant, le nombre d'infections chez les femmes d'origine étrangère reste stable et les infections chez les hommes ayant des rapports avec les hommes (HSH) restent nombreuses (14). Cette maladie est essentiellement transmise par voie sexuelle. Elle suit ainsi les routes formées par le « réseau de contacts sexuels » qui est un type de réseau social, pouvant également être superposé à d'autres réseaux (amical, professionnel). Leur étude a tout d'abord été intuitive et bornée à la recherche de la caractérisation des contacts et des groupes à risques (15). Il est devenu évident que des facteurs de risques propres, liés à la structure du réseau, non personnels, entraient en ligne de compte dans le risque d'acquisition du VIH et dans l'évolution du processus épidémique (16). Un problème crucial est alors la possibilité d'observer ce réseau d'interactions. Actuellement, la mesure globale de réseaux d'interactions est rare car techniquement difficile (17). On dispose généralement plutôt de données 'égocentriques' où des personnes interrogées décrivent les personnes avec lesquelles elles sont en interaction, mais l'information sur les relations entre '*alter'* est manquante. La connaissance du réseau sexuel est ainsi devenue un enjeu majeur dans l'étude de la transmission du VIH dans les différentes populations et également dans l'évaluation des moyens de contrôle de l'épidémie. Depuis l'avènement de la Prophylaxie de Préexposition contre le VIH (PrEP) les modèles mathématiques évaluant l'impact des stratégies de prévention prennent en compte des variables de réseaux (18).

Ce travail est divisé en trois parties. La première définit la notion de réseau, son intérêt en épidémiologie et dresse le panorama bibliographique sur l'utilisation des variables de réseaux dans l'analyse du risque de la transmission du VIH. Bien que les travaux de recherche présentés dans cette thèse ne mettent pas au premier plan l'analyse de jeux de données de réseaux, cette première partie a été jugée indispensable au sens où la modélisation et la simulation se basent toujours sur l'analyse des données réelles. Leur analyse et la recherche bibliographique restent

des prérequis incontournables à la modélisation et à la simulation d'épidémies. Naturellement, les variables de réseaux mises en évidence comme des risques d'acquisition d'une maladie sont intégrées ou testées dans les modèles épidémiques basés sur les réseaux.

La seconde partie détaille l'usage des réseaux dans la simulation de l'évolution de l'épidémie liée au VIH et leur impact sur l'effet des mesures de prévention. Elle se rapporte au premier article publié en lien avec cette thèse.

La troisième partie développe une méthode d'estimation de la prévalence du VIH utilisant la modélisation de réseaux à partir de données d'observation partielle. Il se rapporte au deuxième article écrit en lien avec cette thèse.

Les deux travaux de recherche sont élaborés à partir d'études égocentriques réalisées en France et au Brésil.

1. Réseaux, épidémie et VIH : approche analytique

1.1 Les réseaux

1.1.1 Définitions

Un réseau est défini par un ensemble de nœuds (*vertex* en anglais) et de liens (*edges* en anglais). La présence d'un lien entre deux nœuds définit une relation par laquelle l'échange de messages informatifs est possible (figure 1) (19). Cette définition générale permet l'application des réseaux à différents domaines, de l'informatique aux sciences sociales et à l'épidémiologie. Les nœuds peuvent ainsi représenter des individus, et les liens peuvent représenter toute forme de relation sociale (de pouvoir, d'amitié, commercial, amoureuse, sexuelle, de contact physique). Les messages véhiculés par ces liens peuvent donc être multiples, même pour un même lien, et influencer l'état d'un nœud, dans notre cas, un individu et son état de santé. Ainsi, dans l'étude des maladies infectieuses sexuellement transmissibles au sein de réseaux, les nœuds sont des individus, les liens sont les partenariats sexuels et les messages véhiculés par les liens peuvent correspondre aux micro-organismes portés et potentiellement transmis (gonocoque, chlamydia, tréponème, VIH, hépatites).

L'existence d'un tel réseau permet de décrire l'environnement d'un individu dans plusieurs dimensions qui peuvent être regroupées dans au moins 4 groupes (20):

1) le degré qui représente le nombre de liens d'un individu, possiblement subdivisé selon les types de liens ;

2) les variables qui comparent les caractéristiques des sujets à leurs contacts. Ces variables sont des variables dites *d'homophilie* (différence d'âge, différence de niveau socio-économique…) ;

3) des variables définissant les caractéristiques des *alters* ;

4) les mesures à l'échelle du réseau définissant sa structure : sa taille, sa densité, son nombre de clusters, de diades, de triades, de triangles ;

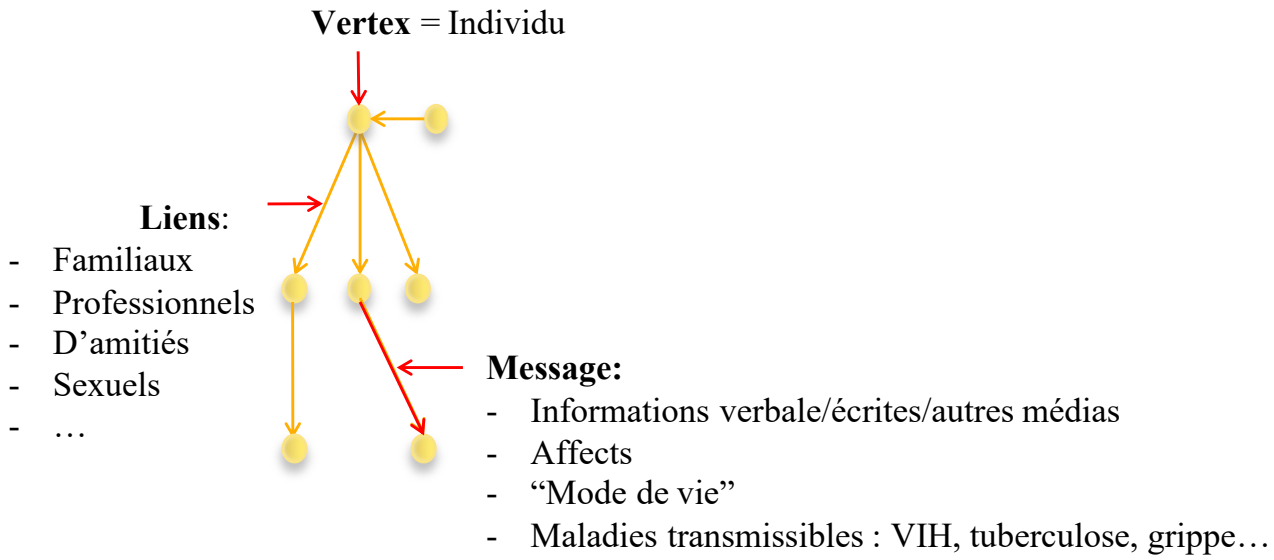

Figure 1 Définition générique d'un réseau et de ses constituants et application à l'épidémiologie. (vertex = nœud)

1.1.2. Les différents types de données de réseaux

On peut rassembler les travaux portant sur les réseaux dans trois grandes familles selon les données utilisées (5). Ils sont représentés dans la figure 2.

Les données socio-centriques : elles sont obtenues par l'observation d'un réseau complet. Elles présentent l'avantage de l'exhaustivité mais ont comme inconvénient majeur la difficulté de collecter des données, coûteuse et complexe. Il existe cependant des jeux de données complets réalisés dans différents travaux de sociologie au sein d'environnements de petite taille pouvant être considérés comme clos, comme les monastères ou les écoles. Dans l'étude des processus épidémiques où on s'intéresse à des réseaux de contacts de grande taille, la collecte

de ce type de données est presque impossible en dehors de communautés bien circonscrites (17).

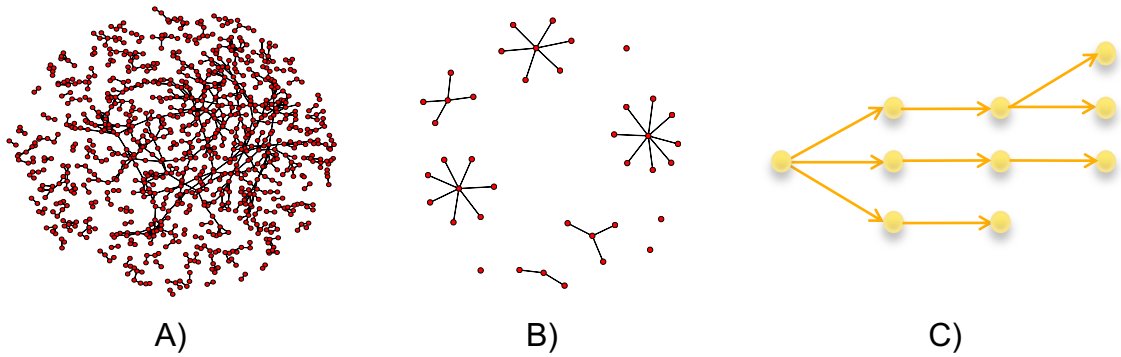

Figure 2 Représentation d'un réseau sociocentrique (A) ; ego centrique (B) et mixte (C).
Le C représente une chaine de recrutement à partir d'un individu. Une partie du réseau est réellement observée mais cette observation est limitée par le nombre d'individus recrutés qui n'est pas égal au nombre réel d'alters. Les relations entre les alters ne sont pas non plus observées.

Les données *ego*-centriques : Elles sont obtenues à partir d'individus (les *egos*) questionnés sur leurs caractéristiques propres, les caractéristiques de leurs partenaires (*alters*) et les liens qui unissent *l'ego* à l*'alter,* voire parfois les liens qui unissent les *alters* entre eux. Ce type de donnée est issu de travaux de recherche dit *ego*-centriques. Ils comportent au moins trois questionnaires à destination de l*'égo* :

- Un questionnaire décrivant *l'ego*, similaire à ceux réalisés dans les études épidémiologiques et comportementales classiques et comportant les données démographiques, socio-culturelles et propres à l'étude.
- Un questionnaire générique relationnel décrivant les *alters* en lien avec *l'ego* au cours d'une période de temps définie. Il en décrit leur nombre, leurs caractéristiques sociodémographiques, quantifie leur nombre par catégorie d'intérêt (par exemple

combien de partenaires sexuels de plus de 30 ans avez-vous eu au cours des 6 derniers mois ?) ou qualifie une variable selon une autre variable (par exemple : utilisez-vous le préservatif avec vos partenaires rencontrés dans un sauna ?).

- Un questionnaire décrivant précisément un nombre *n* d'alters, en général entre 3 et 5, de manière individuelle. Le questionnaire débute en général par une question visant à nommer ou à donner un pseudonyme aux *n* alters qui seront décrits. Puis le questionnaire se personnalise (par exemple : « Quel âge a monsieur X ? »)
- A ces questionnaires s'ajoute parfois une partie tentant de décrire les liens entre les alters (par exemple : Est-ce que monsieur X connaissait monsieur Y ?)

A partir de ces données, il est possible d'élaborer des variables de réseaux à intégrer dans le risque personnel d'acquérir une pathologie. Il est également possible d'élaborer des réseaux égocentriques répondant aux statistiques construites à partir des réponses aux différentes parties du questionnaire (cf. infra) (21). Cet outil présente l'avantage de ne pas nécessiter de questionner l'ensemble des nœuds du réseau et est donc moins couteux en temps. Les inconvénients sont l'absence d'information sur les structures complexes du réseau comme le nombre de boucles, de triangles, etc. et la problématique d'avoir à se fier à une description indirecte des *alters* par *l'ego*.

Les données mixtes. On parle de données de réseaux mixtes lorsqu'une partie du réseau est observée et que des informations complémentaires sur les alters sont demandées aux ego recrutés. Ces données sont obtenues par les processus de recrutement par les pairs, en chaine. Dans ces études, il est demandé aux personnes recrutées de recruter dans l'étude leurs contacts. La chaine de recrutement représente donc une partie du réseau de contact réel auquel il manque des alters (qui peuvent être décrits dans un questionnaire ego-centrique) et des liens entre les personnes recrutées. Il n'est en effet pas possible de demander aux recrutés s'ils connaissent

d'autres personnes recrutées en dehors de leurs recruteurs et des individus qu'ils ont eux-mêmes recrutés ; sauf dans une certaine mesure si le design de l'étude autorise qu'une personne puisse être recrutée plusieurs fois. Un cas particulier de ce type de données est celui issu des enquêtes épidémiologiques de type Respondent Driven Sampling (RDS), ou enquête dont l'échantillonnage est dirigé par les répondants (cf. infra) (22).

1.2. Utilisation des données de réseaux en épidémiologie analytique

1.2.1. Généralités

Une partie de l'épidémiologie analytique vise à rechercher les facteurs de risques associés à l'acquisition d'une maladie. Ces facteurs de risques sont souvent dits « personnels » au sens où ils décrivent un comportement, une caractéristique personnelle (génétique, acquise) qui favorise l'apparition d'une maladie. Les facteurs dits « environnementaux » ou non personnels sont également pris en compte notamment dans la large littérature portant sur l'acquisition des cancers. La notion d'interaction entre facteurs de risques personnels et environnementaux est également développée. L'impact de l'environnement social est très largement pris en compte dans l'acquisition des maladies chroniques, qu'elles soient psychiques ou physiques. Cependant, dans la grande majorité des études, ces « variables sociales » (niveau socio-économique, insertion professionnelle) sont des variables définissant l'individu dans la société. Elles résument donc des caractéristiques personnelles construites dans un objectif de classer les individus.

La conjonction des variables sociales et environnementales sont des variables définissant l'individu, non pas par ses caractéristiques propres, mais par les caractéristiques le définissant au sein d'un milieu social. Comme décrit plus haut, le milieu social est un environnement qui peut être considéré comme un réseau de contacts dans lequel l'individu évolue. On peut donc supposer que les liens et les variables aboutissant à la création de ses liens sont des variables ayant une influence sur l'individu. Ces variables issues du réseau peuvent donc avoir une influence sur l'acquisition d'une maladie.

Les modèles les plus simples sont utilisés intuitivement en prenant par exemple des « figures » du réseau social et l'impact de leurs caractéristiques sur l'individu étudié définissant la « contagion sociale » (23–25). Les variables d'individus extérieurs sont additionnées aux variables personnelles pour évaluer un risque d'acquisition d'une maladie. La prise en compte de plusieurs individus formant le réseau voire même leurs interactions entre eux est difficile de par la difficulté de les mesurer (problématique de la visualisation des liens de manière exhaustive) et de par la notion même d'interaction, difficile à intégrer dans un modèle d'estimation du risque (26).

1.2.2. Évaluation du risque d'acquisition du VIH et variables de réseaux

1.2.2.1. Généralités

Très tôt dans l'épidémie liée au VIH, les chercheurs se sont intéressés aux sous -groupes plus ou moins touchés par cette infection, leurs caractéristiques sociales mais aussi la structure des réseaux de contacts sexuels (2,27). Le fait que l'épidémie ne se répande pas de la même manière chez les hommes ayant des rapports sexuels avec des hommes (HSH) que chez les hétérosexuels ou qu'il existe des différences de prévalence chez les hétérosexuels blancs, noirs ou hispaniques aux États-Unis à facteurs de risques personnels similaires, a rapidement soulevé des questions

portant sur des variables non habituellement prises en compte dans le risque de transmission du VIH (28) (29). Classiquement, le risque d'acquisition du VIH est associé au nombre de partenaires (qui est déjà une variable de réseau), à l'usage d'outils de prévention (préservatif, et maintenant, prophylaxie de pré exposition ou PreP), aux types de rapports sexuels (anaux ou vaginaux, réceptifs ou insertifs), à l'usage de drogue et d'alcool. Ces risques sont dits « personnels » au sens où ils définissent les pratiques de l'individu. La notion de risque non personnel ou lié au réseau a été introduite par l'intégration des types de partenaires (réguliers, occasionnels), de la cinétique des relations entre elles (plusieurs partenaires simultanés, partenaires séquentiels) et des variables construites par la conjonction de variables d'un individu (*l'ego*) et de ses partenaires (les *alters*). On peut citer les différences d'âge, de niveau socio-économique, et éducationnelles (30–33,33,34).

L'identification des variables de réseaux et leur impact dans le risque d'acquisition du VIH, mais également dans le risque de développer des conduites à risque d'infection, a été développée par plusieurs équipes s'intéressant à différentes populations clefs. La méthodologie classiquement utilisée était de réaliser des études égocentriques sur ces populations et d'estimer l'importance des variables individuelles et liées aux réseaux de contacts sexuels dans l'acquisition du VIH.

Ces travaux, essentiellement nord-américains sont élaborés schématiquement de deux manières différentes : 1) L'étude d'un facteur de risque en particulier et 2) la comparaison de groupes présentant ou non une infection ou un risque d'infection en intégrant les variables personnelles et les variables de réseaux.

1.2.2.2. Description des variables qui peuvent avoir un impact sur la transmission de l'infection.

Un premier groupe de travaux porte sur les variables caractérisant le lien entre les *egos* et les *alters*. La différenciation des relations régulières et occasionnelles est historique et fondamentale dans l'évaluation du risque de transmission du VIH par voie sexuelle (35–43). Il s'agit déjà d'une variable de réseau qui définit les types de liens sur lesquels la transmission du virus peut se produire. Les réseaux de partenaires réguliers et occasionnels forment des réseaux de contacts « superposés » ayant des caractéristiques dynamiques et structurelles différentes qui jouent un rôle dans les caractéristiques de l'épidémie (44). Les variables d'homophilie par âge ou par groupe racial ont été étudiées individuellement dans un objectif de compréhension des disparités ou de changement des valeurs d'incidences de la maladie entre groupes d'individus aux caractéristiques sus-citées identiques (45) (46–48).

Ensuite, des variables structurant le réseau de contacts sexuels et pouvant être associées à des facteurs de risques personnels ont également été étudiées dans leur dimension « réseau », notamment les variables d'organisation sociale comme les lieux de rencontre fréquentés. La variable « lieu de rencontre » est par définition une variable d'homophilie forte (49–52). Son impact sur l'épidémie du VIH est de plusieurs ordres. Ces lieux peuvent être considérés comme des proxy résumant des pratiques sexuelles réalisées avec les personnes rencontrées dans ces lieux, mais ils peuvent également favoriser l'interconnexion facilitant la transmission du VIH et d'autres infections sexuellement transmissibles (53). Étudier l'impact de la fréquentation de ces lieux présente un intérêt double car il permet d'évaluer une variable de réseau facile d'accès mais il s'agit également d'une des rares variables de réseaux modifiables ou utilisables à des fins de prévention.

Les réseaux ont également une structure temporelle qui influence le déroulement du processus épidémique. La caractéristique étudiée de manière extensive est la notion de partenaires sexuels simultanés (concurrency) (54). Cette variable répond maintenant à une définition validée par l'organisation mondiale de la santé tant elle semble avoir un rôle majeur dans l'épidémie en Afrique subsaharienne mais également dans les pays occidentaux au sein de populations HSH (55,56). L'impact de cette variable s'explique par le fait, qu'à la différence de partenaires sexuels séquentiels, un individu en lien avec un alter infecté ayant ce type de relation maintient une probabilité non nulle d'être infectée tout le long de la relation. Ainsi, à nombre de partenaires identique sur une même période de temps et à nombre de rapports constant, le risque de transmission au sein du réseau est plus important en cas de partenaires simultanés qu'en cas de monogamie séquentielle (figure 3). La notion de partenaires simultanés est donc un facteur de risque d'acquisition du VIH mais également un facteur de risque de transmission majeur au sein du réseau.

Enfin, la variable du nombre de partenaires sexuels est connue comme étant un facteur personnel majeur de transmission du VIH. Elle peut également être vue comme la première variable constituant le réseau et définit le « degré » de lien. La notion de réseau interpelle sur l'interprétation de cette variable comme une densité qui est définie par le nombre de liens rapportés au nombre d'individus du réseau. Cette première variable structurante peut ainsi avoir une importance plus ou moins forte dans un réseau selon les autres caractéristiques de celui-ci (figure 3). Ainsi, la notion du nombre seuil de partenaires définissant le risque de transmission fait intervenir l'environnement dans lequel ces liens se créent et peut avoir un impact plus ou moins fort notamment en présence de partenaires simultanés plutôt qu'en présence de « partenaires séquentiels » (57).

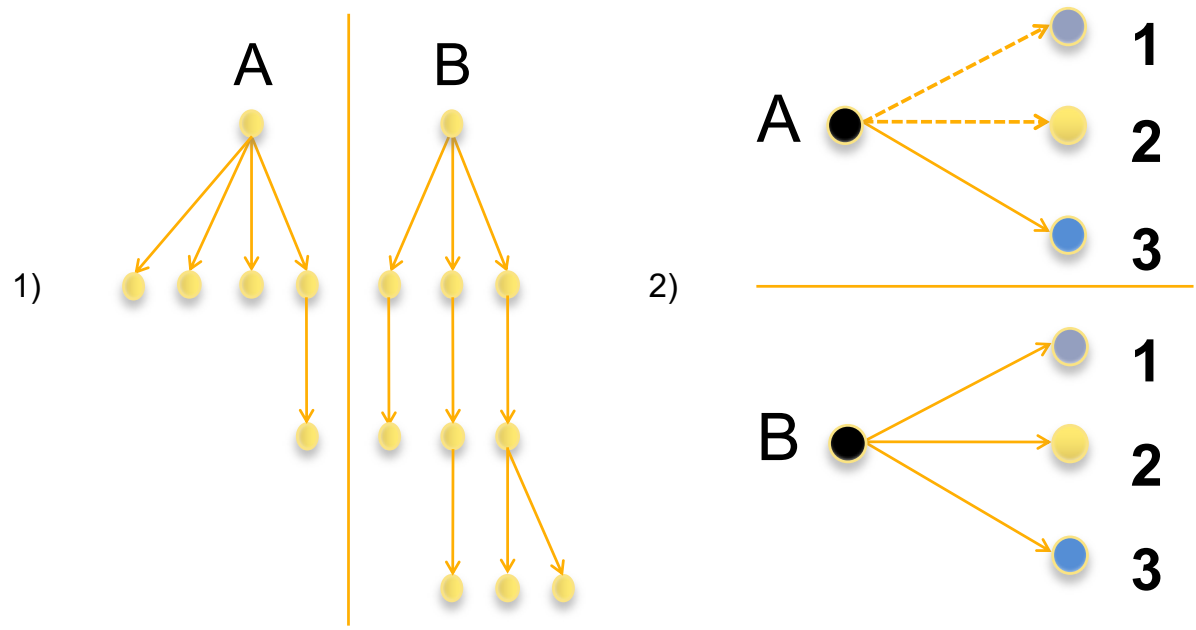

Figure 3 Variables de réseaux ayant un impact dans la transmission.

1) Exemple de réseau où le degré de l'ego ne présage pas de la taille de son réseau et donc du nombre de personnes susceptibles d'être infectées. 2) Visualisation de liens séquentiels (haut) et simultanés (bas). Le nœud 2 est infecté et les autres nœuds sont susceptibles d'être infectés. Dans la situation de liens séquentiels, seul le nœud A peut être infecté, puis le nœud 3 si A a été infecté. Le nœud 1 ne peut pas être infecté. En présence de liens simultanés, tant que le lien entre 2 et B est possible, le risque d'infection des nœuds 1 et 3 est non nul.

A titre d'illustration, et en guise d'introduction à la simulation d'épidémie au sein de réseaux, la table 1 et la figure 4 représentent les résultats d'une étude cas-témoins virtuelle simulée dans un réseau de contact où tous les individus ont la même probabilité de présenter des liens et présentent des facteurs de risques personnels d'acquisition identiques (non publié). Ce travail repose sur des réseaux théoriques de 10 000 individus dans lesquels seul le nombre de liens à chaque instant est défini. A chaque pas de temps, des liens disparaissent et se construisent mais leur nombre reste constant. Une épidémie est simulée à partir de 10 individus (nœuds) infectés au sein de ce réseau. A chaque pas de temps, une infection peut être transmise selon une probabilité définie par un nœud infecté vers les nœuds non infectés liés à lui. Les données sont récoltées à chaque pas de temps. Les vertex nouvellement infectés sont comparés sur leurs variables de réseaux à trois personnes non infectées au temps de l'infection. Plus précisément,

les réseaux de liens cumulés des 90 derniers pas de temps centrés sur les cas et les témoins sont extraits jusqu'à l'ordre 3 (réseaux comprenant l'ensemble des *alters* liés à *l'ego* directement et indirectement jusqu'à deux individus intermédiaires et sur les 90 pas de temps antérieurs) et comparés. Les résultats montrent que bien que significatif, le nombre de liens direct d'ordre 1 n'est pas très différent entre les personnes infectées et non infectées (4 vs 3 partenaires sexuels au cours des 90 derniers pas de temps) mais les autres variables de réseaux sont différentes. Ainsi, à facteur de risque personnel similaire, le risque d'être infecté est dépendant de la structure du réseau de contact.

Variables	Estimation	sd	p-value
Degré	1.33	0.11	0.0019
Taille du réseau	1.28	0.09	0.0008
Nb de liens totaux	1.02	0.01	0.061
Nb alter ordre 2	1.67	0.44	0.0018
Partenaires simultanés	3.55	1.08	0.001

Tableau 1 Résultat de la régression comparant les données de réseaux des individus infectés aux individus non infectés.

Les données de réseaux comparées comportent des variables caractérisant les réseaux jusqu'à l'ordre 2, c'est à dire s'étendant jusqu'aux contacts des alters.

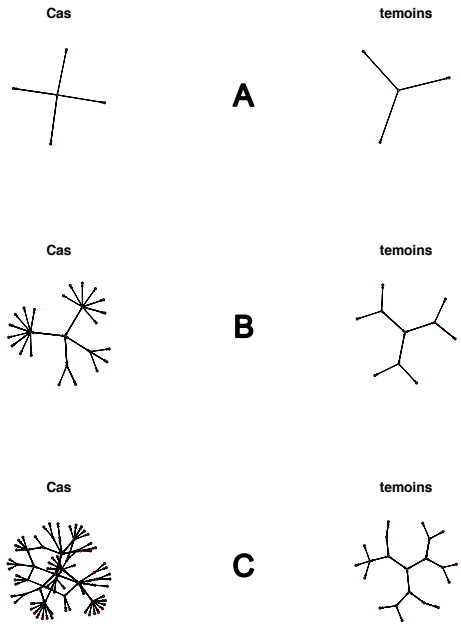

Figure 4 Représentation d'un réseau représentatif d'un individu infecté (cas) et d'un individu non infecté (témoins).

A représente les informations récupérées dans une enquête épidémiologique classique (nombre de partenaires de l'égo) ; B et C représentent les informations accessibles dans des études de réseaux s'intéressant aux caractéristiques des partenaires (B) et des partenaires de partenaires (C) (liens d'ordre 2 et 3, respectivement).

1.2.2.3. Prise en compte des risques liés aux réseaux dans l'acquisition d'une maladie ou d'un comportement à risque

Des travaux s'intéressant aux variables connues comme potentiellement d'intérêt dans la transmission du VIH ont intégré la notion de risques liés aux réseaux. Pour exemple, Ils portent sur les différences dans les réseaux sociaux, sexuels ou amicaux, entourant les personnes vivant avec le VIH (PVVIH) nouvellement diagnostiquées comparativement aux personnes non infectées d'un même milieu (58) ou le risque d'être infecté par le VIH chez les HSH selon leur « groupe raciaux » aux États-Unis d'Amérique (59). D'autres travaux ont évalué l'importance du réseau de contact dans l'acquisition de pratiques à risque de transmission sexuelle (60). Ces travaux intègrent la taille du réseau de contact, l'homophilie selon des variables

sociodémographiques, la notion de partenaires simultanés. Carl Latkin et son équipe ont développé l'analyse des données égocentriques dans plusieurs populations clefs de l'épidémie liée au VIH. Leurs travaux sur les réseaux de contact portent essentiellement sur les HSH américains, la population noire américaine HSH et les usagers de drogue intraveineuse (53,59,61,62). Ils intègrent les facteurs de risque non personnels pour expliquer le risque d'acquisition du VIH mais également pour décrire des comportements à risque ou le développement de pathologies autre comme la dépression dans ces mêmes populations (63). Les analyses des réseaux de contact des usagers de drogue sont plus développées par ces équipes et dans la littérature que celles portant sur d'autres populations clefs, probablement du fait du caractère plus circonscrit et individualisable du réseau permettant la réalisation d'enquêtes quasi-sociocentriques (64). Ce réseau de contact est pourtant complexe à analyser car il intègre un réseau de contacts amicaux, de partage de drogue et de seringue mais également un réseau de contacts sexuels. Des travaux recoupent ces différents réseaux dans l'évaluation du risque d'infection par le VIH des usagers de drogues. Ils décrivent que les membres d'un réseau de contact d'usagers de drogues séropositifs présentent un nombre d'individus séropositifs plus important que le réseau de contact d'usagers de drogues séronégatifs. Ceci s'explique par la transmission directe par voie sexuelle et intraveineuse de l'infection mais également par le fait que les personnes ayant des pratiques à risque identiques se regroupent (homophile par le risque individuel) (65). Les autres variables de réseaux mises en évidence chez les usagers de drogues sont des structures de réseaux particulières permettant la transmission comme un minimum de deux liens à risque pour chaque individu (66), la proximité dans le réseau d'un individu séropositif (67), la non assortativité par tranche d'âge (68) ou la superposition de réseaux de contacts amicaux et sexuels (69).

L'étude des variables de réseaux de la transmission du VIH dans d'autres populations clefs démontre l'impact majeur des partenaires simultanés chez les HSH et chez les hétérosexuels en

Afrique subsaharienne (70–72)(73), de l'assortativité par groupes raciaux aux États-Unis et par d'autres variables comme l'âge ou le niveau socio-économique (46). Dans la population HSH, l'évaluation du risque d'infection par le VIH a été mesurée en intégrant les variables de réseaux mais également la fiabilité des connaissances portant sur les alters (74).

1.2.2.4. De l'analyse des données de réseau vers la modélisation de réseau

La littérature portant sur l'analyse des données de réseaux pour expliquer le risque d'infection par le VIH ou d'autres infections sexuellement transmissibles est ainsi devenue pléthorique. Elle confirme que les variables de réseaux peuvent être associées au risque d'acquérir le VIH ou une conduite à risque. Cependant, ces variables ne sont, pour la plupart, pas ou difficilement modifiables. Il est en effet difficile d'intervenir sur l'homophilie par groupe social ou par âge à des fins de prévention. Des campagnes d'information sur les risques liés aux partenaires simultanés ont été essayées mais la modification de telles variables n'est pas aisée et peut se heurter aux pratiques socio-culturelles des populations ciblées. Les caractéristiques de réseaux peuvent en revanche être utilisées pour favoriser la transmission d'un message de prévention (53). L'intérêt de ces variables réside aussi dans leur capacité à mieux définir une population à risque d'infection. Enfin, le fait que ces variables soient retrouvées associées au risque d'infection par le VIH à l'échelle individuelle incite à évaluer leur importance dans l'évolution de l'épidémie et sur leur impact sur l'efficacité des mesures de prévention.

Ceci peut être étudié par la simulation de processus épidémiques au sein de réseaux.

2. Impact des réseaux dans les processus épidémiques et leur prévention : modélisation de réseaux et simulation d'épidémie

2.1. Généralités

L'évolution d'une épidémie est résumée en grande part par le ratio de reproduction (75). Cette valeur, notée R, correspond au nombre de personnes directement infectées par un individu infecté. Lorsque l'on se place dans une population infinie totalement susceptible, on note la valeur correspondante R_0, qui correspond alors au ratio de reproduction basique. Ce paramètre, remis en lumière dans les travaux de Klaus Dietz (76) pour les modèles épidémiques en général, suivant ceux de George Mac Donalds (77) dans le modèle de transmission d'une maladie vectorielle. Dans le cas du modèle SIR (**S**usceptible **I**nfectious **R**ecovered) standard d'Anderson et May (1), le ratio de reproduction s'exprime classiquement comme le produit des 3 paramètres:

-ß : la probabilité de transmission par contact,

-D : la durée de la période infectieuse.

-c : le taux de contact entre les personnes,

Ces trois composantes ont des valeurs différentes selon les pathologies et les populations atteintes. Plus encore, pour l'infection par le VIH, la probabilité de transmission par contact dépend par exemple du stade de l'infection, des coinfections, et de l'efficacité des traitements antirétroviraux (78,79)(80). La durée infectieuse est variable d'un individu à l'autre, selon la connaissance de son statut, les traitements et les comportements adoptés.

Cette expression simple de R_0 repose sur l'hypothèse d'homogénéité des contacts, c'est à dire que tous les individus sont en contact avec la même probabilité. Les premiers modèles de

transmission de l'infection par le VIH acceptaient les hypothèses simples d'homogénéité de contact (78). Cependant cette description ne permet pas de représenter l'organisation sociale des relations interindividuelles et inter populationnelles. Dans les années 80 se sont développés des outils de modélisation de processus épidémiques prenant en compte l'hétérogénéité des contacts et des mélanges des populations (81). Pour l'étude de l'infection liée au VIH, la modélisation du processus épidémique a pris en compte des caractéristiques réseaux : L'assortativité selon des caractéristiques individuelles (homophilie), pour des groupes d'individus et selon les types de relations (82) (15,83) (84)(85–88).

Les variables de réseaux utilisées en modélisation de la transmission du VIH sont essentiellement celles trouvées comme étant significativement associées au risque d'acquérir le VIH. Les principales sont les suivantes :

- **Le nombre de liens au sein du réseau.** Plus les individus sont connectés, plus la probabilité qu'un lien entre une personne infectée et une personne susceptible de l'être est grand. Cette valeur s'exprime en valeur absolue ou en densité calculée par le rapport du nombre de liens présents sur le nombre de liens possibles (57)(89). Le type de distribution des degrés qui suit habituellement une loi de puissance entraine également des modifications de la transmission de l'infection (90).
- **L'homophilie.** Il s'agit de la capacité d'un individu à créer des liens avec des individus ayant la même caractéristique. Les variables d'homophilie les plus étudiées sont le sexe, l'âge, les catégories socio-professionnelles, l'ethnicité qui déterminent la capacité d'un individu à créer des liens avec un individu de même sexe, âge, catégorie socio-professionnelle et ethnicité, respectivement. L'effet de ce type de variable dans le processus épidémique est complexe à déterminer a priori. En effet, une homophilie importante peut permettre de circonscrire l'épidémie à un groupe d'individus, et donc

en limiter l'extension, mais elle peut également la circonscrire à un groupe qui sera plus rapidement infecté (91–93).

- **Les liens simultanés (concurrency).** Cette variable introduit la notion de temporalité dans la structure d'un réseau et nécessite ainsi l'élaboration de modèles dynamiques en vue d'en apprécier l'impact sur le processus épidémique. Sa définition est la capacité d'avoir à un temps donné un lien avec un individu X puis avec un individu Y pour avoir à nouveau un lien avec l'individu X au temps suivant. Cette variable a été énormément discutée dans la littérature. Il est maintenant démontré par la modélisation que cette caractéristique facilite la propagation d'une infection comme celle liée au VIH, que ce soit dans les populations ayant des rapports hétérosexuels et dans les populations d'HSH (9,94,95). L'interrelation entre cette variable et les autres propres aux réseaux est importante, notamment la densité de lien (96), rend l'étude de son impact difficile à individualiser.
- **La durée d'existence des liens.** La transmission d'une infection par un lien ne peut se réaliser que pendant la période durant laquelle ce lien existe. Ainsi, plus un lien existe longtemps entre un individu infecté et un individu susceptible de l'être, plus le risque de transmission est grand, sous réserve que la probabilité de transmission reste non nulle au cours du temps. Les modèles de simulation d'épidémie dans les réseaux utilisent des durées d'existence de liens différents selon les types de liens, réguliers ou occasionnels. Les modèles étant de plus en plus réalistes et prenant en compte des probabilités de contacts sexuels différentes selon les types de partenaires, l'impact de la durée de l'existence du lien sur le risque de transmission devient moins évidente et dépend des autres variables de réseaux, notamment de la notion de partenaires simultanés (29,44,97,98).

2.2. Les outils d'élaboration de la simulation d'épidémie au sein de réseaux

2.2.1 Généralités

L'étude des processus épidémiques repose sur l'observation des données récoltées, leur analyse permettant d'en expliquer l'origine, et la simulation permettant de tenter de prédire son évolution naturelle ou l'impact d'interventions. La simulation repose sur les données et leur analyse afin d'émettre des hypothèses permettant la construction du modèle épidémique réaliste. Idéalement, les modélisations doivent pouvoir reproduire les données réelles de l'épidémie.

Dans le cadre de l'étude de l'impact des variables de réseaux dans une épidémie, les travaux de simulations reposent sur deux éléments :

- la création de réseaux ou graphes comportant les caractéristiques d'intérêt. Il repose sur l'analyse d'un jeu de données égocentriques ou sociocentriques. Des réseaux théoriques comportant des caractéristiques d'intérêt peuvent également être développés.
- Un modèle de transmission définissant la probabilité de transmission entre un individu infecté et un individu susceptible d'être infecté au sein de ce réseau à chaque pas de temps. L'influence de chaque variable choisie dans le modèle de transmission est définie par des données issues de la littérature scientifique.

2.2.2. Élaboration de graphes

L'analyse des réseaux et leur impact sur la transmission nécessite leur visualisation ou tout du moins leur élaboration. La théorie des graphes permet cela en définissant les nœuds et les liens liant les nœuds et en permettant leur visualisation sous forme de points et de traits. Elle permet de décrire et donc de définir les réseaux par leurs caractéristiques en termes de taille, d'orientation, de densité de liens et de structures particulières. Elle permet également de définir la nature du lien. Leur développement est traditionnellement associé à Euler au XVIIIe siècle qui étudia la problématique de traverser l'ensemble des ponts de la ville de Königsberg, en partant d'un quartier quelconque de la ville, sans passer deux fois par le même chemin. Au vingtième siècle, les graphes deviennent une part des mathématiques. Leur élaboration est rendue possible par des théorèmes et des procédés algorithmiques (99).

A partir des années 1980, l'utilisation de graphes pour étudier la transmission des messages s'est développée et a été appliquée à la transmission d'infections. La modélisation de réseau à des fins d'études d'épidémies a nécessité de formaliser l'élaboration des réseaux afin de rendre reproductible les expériences de simulation et d'évaluer l'impact de la modification de ces réseaux dans le processus épidémique. Les outils utilisés pour réaliser des réseaux simples sont les graphes de configuration, définis par les degrés de chaque nœud (100) et les modèles dans lesquels la distribution des degrés suit une loi prédéfinie comme les modèles de Erdős-Rényi ou de Barabasi (101,102). Dans le cadre de situations nécessitant d'introduire plusieurs autres variables dans la construction des réseaux, les graphes aléatoires exponentiels (exponential random graph : ERGM) présentent une solution mathématique performante (103).

2.2.3. Les ERGMs : Un outil mathématique de modélisation de réseau

L'ensemble des variables de réseaux décrites plus haut sont intégrables au sein de modèles mathématiques permettant d'élaborer des réseaux de contacts répondant aux caractéristiques proposées. Elles peuvent également être estimées à partir d'un réseau observé. L'outil développé dans ces deux objectifs est l'exponential random graph (ERGM) dans sa composante statique et dynamique (104,105).

L'ERGMs est un outil mathématique, initialement appelé « p model », issu des sciences sociales . L'ERGM définit la probabilité de l'existence d'un lien entre deux nœuds du réseau selon des caractéristiques globales du réseau, du type de lien et des caractéristiques des individus qui sont impliqués dans le lien. Le modèle utilisé est dit « auto-logistique », c'est-à-dire qu'une régression logistique modélise la probabilité d'existence de chaque lien, en fonction des autres liens et des caractéristiques associées aux nœuds. Dans cette formulation, les paramètres du modèle correspondent à des odds-ratio mesurant le « risque » que le lien existe ou non en présence de la caractéristique donnée (106–108).

La formulation d'un ERGM repose sur un vecteur de statistiques locales et globales calculées sur un réseau Y, noté g(y), qui peut comprendre par exemple le nombre de liens, le nombre de nœuds ayant un certain degré, des caractéristiques des nœuds et l'homophilie selon ses caractéristiques, des caractéristiques des liens. Chaque statistique entre dans l'expression du modèle sous la forme de la différence des valeurs lorsqu'un lien est présent ou non. Plus spécifiquement, le lien entre les nœuds i et j, noté $Y_{ij}=1$ s'il est présent, est spécifié conditionnellement au reste du réseau (Y_{ij}^{c}) selon :

$$\text{Logit}\,[P_{\theta,Y}(Y_{ij}=1|Y_{ij}^{c}=y^{c}_{ij})]=\theta^{T}\delta_{g}(y)_{ij}$$

Où $\delta_{g}(y)_{ij}=g(y_{ij}^{+})-g(y_{ij}^{-})$ avec :

$g(y_{ij}^{+})$ = valeur de g(y) lorsqu'il existe un lien entre i et j

$g(y_{ij}^{-})$ = valeur de g(y) lorsqu'il n'existe pas de lien entre i et j

La bibliothèque statnet du logiciel R (109) permet la création, la modélisation et la simulation de réseaux à partir de données socio-centriques ou égocentriques à l'aide des ERGM.

Cet outil peut donc être utilisé de deux manières. Premièrement et à partir d'un réseau observé (données sociocentriques), il permet de déterminer les variables d'importances définissant ce réseau. Cette utilisation est rare dans l'étude des processus épidémiques car les réseaux de contacts ne sont que rarement observés. Deuxièmement, ils permettent de créer des réseaux ayant des caractéristiques préalablement définis à partir de statistiques voulues (ou connues) pour chaque paramètre d'intérêt. Cette utilisation permet, à partir de données *ego*-centriques de créer des réseaux de contact. Il permet également de réaliser des réseaux théoriques répondant aux caractéristiques voulues. Ces deux dernières utilisations sont les plus fréquentes dans l'étude de la propagation d'une épidémie.

2.2.4. L'ERGM et sa composante dynamique (105,110,111)

Les ERGM décrivent des réseaux dit statiques qui peuvent être un réseau de liens cumulés sur une période de temps plus ou moins longue. Les liens sociaux et plus spécifiquement pour notre travail, les contacts sexuels évoluent avec le temps. Il est donc nécessaire de développer des modèles autorisant la création de suite de réseaux dans lesquels les caractéristiques des individus et des liens évoluent avec le temps. Ce processus est de type markovien au sens ou le réseau au temps t dépend de l'état du réseau au temps t-1. L'outil ERGM a donc bénéficié d'un développement dynamique, permettant l'élaboration d'une suite de réseau en temps discret (discret-time ERGM ou DT_ERGM).

Plus précisément, les modifications apportées à chaque unité de temps dépendent des paramètres dynamiques semblables à des paramètres définissant l'incidence des modifications

entre chaque temps. Nous avons adopté la spécification 'séparable' décrite par Krivitsky et al, qui spécifie indépendamment l'addition de nouveaux liens et la disparition de liens existants sous la forme :

$$P(Y_t \mid Y_{t-1}) = P_{\theta+,\gamma}(Y^+=y^+|Y_{t-1}) \times P_{\theta^-,\gamma}(Y^-=y^-|Y_{t-1})$$

Les deux parties de l'équation sont définies par :

$$P_{\theta+,\gamma}(Y^+=y^+|Y^{t-1}) \propto \exp[\theta^{+T}\delta_{g+}(y^+)]$$

C'est la probabilité de passer de l'état y^+ à partir de l'état Y_{t-1}. Les états y^+ autorisés sont ceux qui résultent de l'addition de liens à Y^{t-1}. Les $\delta_{g+}(y^+)$ correspondent aux différences des statistiques entre ces deux états.

$$P_{\theta-,\gamma}(Y^-=y^-|Y^{t-1}) \propto \exp[\theta^{-T}\delta_{g-}(y^-)]$$

Chaque pas de temps décrit un réseau de liens cumulé sur une période de temps prédéfinie. La durée de cette période dépend des connaissances de l'impact des liens cumulés sur le processus épidémique. Dans le contexte de la transmission des infections sexuellement transmissibles et devant la notion de partenaires simultanés qui semble d'importance, le pas de temps d'une journée parait raisonnable.

2.2.5. Les modèles de transmission du VIH

Une fois les réseaux de contacts définis, l'élaboration du processus épidémique en leur sein nécessite de déterminer la probabilité de transmission du VIH d'un individu infecté vers un individu susceptible.

La probabilité d'infection par le VIH est liée à la charge virale de la personne infecté et à l'éventuelle présence d'une mesure de prévention pendant le contact et au type de contact (rapport vaginal, anal, incertif ou non).

L'évolution de la charge virale sans traitement au cours de la maladie a été réalisée sur des données de cohortes à partir desquels plusieurs modèles plus ou moins complexes ont été développés. Les plus connus ont tenté d'évaluer la probabilité de transmission du virus V.I.H chez les couples serodiscordants hétérosexuels en fonction du nombre d'actes sexuels et/ou du stade de l'infection (primo-infection, phase latente, stade SIDA) à partir des données de Wawer (112). Les modèles de Pinkerton (113), Wawer et Abu-Raddad (114) prennent en compte la variabilité du nombre de relation sexuels selon les phases de la maladie. Cette probabilité de transmission sur une durée ß se formalise ainsi :

$$\beta = 1\text{-}(1\text{-}\beta_a)^a$$

où a est le nombre d'actes sexuels par période , β_a la probabilité de transmission par acte sexuel. Le modèle de Hollingworth (115) décrit une probabilité instantanée de transmission β en personne-année. Il ne dépend pas de la variabilité du nombre de rapports selon le stade de la maladie mais décrit trois phases, primo infection, séropositivité chronique, stade SIDA pour lesquelles la probabilité de transmission diffère.

Les modèles plus récents utilisent cette valeur converti en probabilité par pas de temps (116), β_t grâce à l'équation :

$$B_t = 1\text{–}\exp(\text{–}\beta/\text{t}).$$

Où t est la fraction de la période β utilisée comme pas de temps de la simulation.

Cette probabilité β se décline maintenant en prenant en compte les interventions dont la première est la mise en route d'un traitement permettant la baisse de la charge virale, jusqu'à rendre les patients non infectant. L'usage du préservatif, de la PreP par le partenaire, l'observance médicamenteuse et différents profils évolutifs de l'infection sont également pris en compte. Ainsi, La valeur de β devient une variable individuelle calculée à chaque pas de temps selon les paramètres suscités qui prennent en compte les caractéristiques de l'égo infectés mais également de ses alters (98).

2.3 Modélisation des réseaux de contacts sexuels afin d'expliquer l'évolution de l'épidémie liée au VIH et impact des mesures de prévention

2.3.1 Généralités

Les travaux de modélisations de réseaux de contacts sexuels se sont tout d'abord intéressés à l'importance des mélanges ou de l'homophilie entre groupes de patients (85–88). Dans ces articles, plusieurs réseaux théoriques présentant des structures différentes sont développés. Les épidémies sont lancées dans chaque réseau à partir d'une ou plusieurs sources et comparées ensuite en termes de nombre de cas, d'incidence globale et selon les groupes définis. Ces modèles non réalistes, voire théoriques gardent de leur importance car ils permettent de démontrer l'impact des variables de réseaux dans la transmission, mais également d'évaluer à partir de quelle valeur de densité, d'homophilie ou de liens concurrents la variable de réseau est d'intérêt. Ils permettent également d'estimer selon les formes de distribution des degrés l'effet sur la transmission. Peu de travaux s'intéressent à la conjonction de plusieurs variables de réseaux de manière simultanée.

Le caractère initialement très discuté de l'importance de la notion de partenaires simultanés (concurrency) a permis le développement de plusieurs modèles épidémiques prenant en compte des variables personnelles et de réseaux (9,97,117). Depuis 10 ans, les travaux intègrent plusieurs variables d'homophilie déterminant le réseau, rendant les modèles plus complexes et

plus réalistes. La structure du réseau n'est plus discutée, elle est acquise. Le modèle de réseau est validé sur la littérature et des hypothèses d'interventions sont réalisées. Ces modèles sont admirables de par leur complexité. En revanche, ils manquent de justification dans le choix des variables utilisées comme structurantes du réseau. Bien qu'isolément démontrées comme d'importance, les variables d'homophilie peuvent être d'intérêt moindre lorsqu'elles sont additionnées, à l'image de variable significative en univarié qui ne le sont plus en analyse multivariée. Ces modèles présentent en revanche une importance capitale dans l'évaluation des interventions comme la prophylaxie de préexposition (PreP) et de stratégies de prévention combinée (18,118). Actuellement les prévisions d'évolution de l'épidémie liée au VIH aux états Unis réalisées par le Center for Disease Control and Prevention utilise ce type d'outils basé sur les ERGM et un modèle de transmission individu centré.

2.3.2. Modélisation des réseaux de contacts sexuels dans une population HSH française, impact de sa structure sur l'efficacité des mesures de prévention

2.3.2.1 introduction

La modélisation de l'épidémie liée au VIH en utilisant les réseaux en France est débutante. Ceci s'explique probablement par le peu de données ego- ou socio-centriques détaillées existantes. Il n'y a à ce jour aucune étude française publiée développée spécifiquement pour collecter des données de réseau dans un but de modélisation. Pourtant, l'intérêt est majeur devant les disparités de prévalence du VIH, notamment au sein de la population HSH. Il existe cependant quelques variables de réseaux collectées dans le cadre de travaux épidémiologiques nationaux, notamment dans les études PREVAGAY conduites par Santé Publique France en 2009 et 2015 (119,120).

Ainsi, à partir de ces données, a été proposé de réaliser un travail de simulation de processus épidémiques au sein de réseaux de contacts sexuels « réalistes », c'est à dire proche des données de l'enquête PREVAGAY 2009, et « théorique » dans lesquelles les variables de réseaux sont modifiées en vue d'en évaluer l'importance. Les variables de réseaux décrites dans l'étude, et donc modifiées dans ce travail, sont les lieux de rencontre visités et les densités de liens au sein de ces lieux.

En parallèle, l'apparition de la PreP comme nouvelle mesure de prévention permet de s'interroger sur l'impact de son efficacité populationnelle selon les réseaux. L'hypothèse étant que la diminution du risque individuel d'être infecté défini dans les essais cliniques a un impact populationnel différent selon la structure du réseau.

Ainsi, ce travail repose sur les données de l'étude PREVAGAY 2009 pour l'élaboration des réseaux et sur la littérature pour le développement du modèle épidémique. Son objectif était de déterminer l'impact de variables de réseaux dans la dissémination du VIH mais également dans l'efficacité de la PreP.

Le protocole expérimental a été le suivant :

- Analyse des variables PREVAGAY pouvant être considérées comme des variables de réseaux
- Réalisation de réseaux de contacts dynamiques à l'aide des ERGM:
    - Prenant en compte les variables décrites dans PREVAGAY
    - Modifiant les variables décrites dans PREVAGAY : lieu de rencontre et densité de liens par lieu
- Élaboration d'un modèle épidémique donnant à chaque pas de temps une probabilité de transmission du VIH sur les données de la littérature
- A l'aide de sources infectées prises au hasard, réalisation d'épidémies au sein des réseaux définis (réalistes et modifiés) :

- Sans PreP ;
- Avec des niveaux de couverture PreP différents mais dont la distribution dans la population simulée est faite au hasard ;
- Avec des niveaux de couverture PreP différents mais dont la distribution dans la population simulée est faite dans des populations particulières définies par les lieux de rencontre visités.

Le modèle « réaliste » a été validé par l'obtention de prévalences de l'infection par le VIH dans les différents lieux similaires à celles décrites dans les données PREVAGAY 2019 une fois la prévalence globale réelle atteinte par simulation dans le réseau de contact dit « réaliste » sans la PreP.

Les différents scénarios d'intervention PreP ont ensuite été comparés en termes de diminution du nombre d'infection global et par lieu de rencontre visité, permettant d'estimer le bénéfice direct et indirect de la PreP selon les stratégies de couvertures.

La suite de ce chapitre détaille les données PREVAGAY, le modèle de transmission développé, l'article publié, et enfin une analyse de ce travail.

2.3.2.2. Étude PREVAGAY 2009 et données de réseaux (119)

L'étude PREVAGAY 2009 est une enquête transversale multicentrique réalisée dans les lieux de rencontre parisien fréquenté par les HSH dans le but d'estimer la prévalence du VIH dans cette population et selon les facteurs de risques. Quatorze établissements commerciaux de convivialité gay ont accepté d'y participer : 5 bars sans sexe et 9 backrooms ou saunas et 919 hommes ont été recrutés. Les participants ont bénéficié d'un test de dépistage du VIH et des hépatites B et C. Ils ont également répondu à un auto-questionnaire comportant des volets sociodémographiques, santé dont le recours au dépistage, sexualité. Les attitudes de préventions avec les partenaires occasionnels et/ou stables ont aussi été collectées. Les résultats majeurs de

cette étude portent sur la prévalence élevée du VIH au sein de la communauté HSH visitant ses lieux soit 17,7%, (IC95% [15,3-20,4]. Le questionnaire comportait également une question portant sur les lieux où les partenaires sexuels étaient rencontrés. La prévalence du VIH différait selon les lieux de rencontres : 19% (95%CI 16–23) peur ceux visitant au moins les clubs, 18% (95%CI 15–21) pour ceux visitant au moins les saunas et 26% (95%CI 22–30) pour ceux visitant au moins les backrooms ($p<0.001$).

Les données de cette étude ont donc été à nouveau analysées dans le but de déterminer les fréquentations des différents lieux et les populations visitant un ou plusieurs lieux de rencontre. Cette distribution de lieux visités n'étant pas liée au hasard, développer un modèle de réseaux sur ces données était ainsi d'intérêt.

Le travail réalisé a donc consisté en la réalisation de réseaux de contacts réalistes basés sur la distribution de nombre de partenaires déclarés et sur les lieux fréquentés. Les épidémies développées dans ces réseaux ont été comparées à celles issues des réseaux théoriques dans lesquels les nombres de partenaires et les lieux de rencontre étaient répartis au hasard.

2.3.2.3 Le modèle de transmission

Un modèle de transmission a été développé afin de déterminer la probabilité de transmission du VIH à chaque pas de temps selon le stade de l'infection, le traitement, le type de rapports sexuels et les mesures de préventions déclarées ou supposées. A chaque pas de temps de un jour, le VIH peut être transmis d'un individu à un autre si, à ce moment, un contact sexuel, défini par la présence d'un lien dans le réseau, existe entre un individu infecté par le VIH et un individu non infecté par le VIH. La probabilité d'être infecté lors du contact sexuel est fortement associé à la charge virale de la source, qui dépend de l'efficacité du traitement et de l'évolution de la maladie (avec ou sans traitement). Les paramètres qui modifient la charge virale peuvent changer tous les mois. L'autre variable qui influence la transmission est la prise de mesures de

prévention personnelle, comme la PreP et l'utilisation de préservatifs. Leur influence sur les probabilités de transmission est également détaillée dans ce chapitre.

Ainsi, à chaque pas de temps, la probabilité de transmettre le VIH est calculée pour chaque individu séropositif et en prenant en compte la caractéristique des liens (usage du préservatif ou non) et des alters (statut sérologique, prise de PreP ou non). Les variables introduites dans le modèles sont décrites dans les points A à H de ce chapitre.

A Probabilité d'infection sans intervention

La probabilité d'infection (β) pour chaque relation sexuelle défini par l'équation suivante :

$$\beta = 2{\cdot}45\ log10(V1/V0) * \beta_0 \quad [1] \quad (2{\cdot}45\ (95\%\ CI\ 1.85–3.26)$$

où β_0 est la probabilité d'infection selon le type de relation sexuelle (ici HSH) pour une charge virale de 4.5 log copies/mL (121,122).

La probabilité d'être infectée à chaque contact est:

$$P = 1-(1- \beta)^n \quad [2]$$

où n est le nombre de rapport sexuel moyen par pas de temps avec ce partenaire.

B Intervention

La probabilité d'infection après intervention est :

$$P' = 1-(1- (1-P_{intervention})* \beta)^n \quad [3]$$

où

$$P_{intervention} = P_{preservatif} + (1- P_{preservatif})* P_{prep} \quad [4]$$

Usage du préservatif: $P_{preservatif}$

Son efficacité dépend de sa fréquence d'utilisation (toujours, parfois et jamais comme décrit dans (121,122).

Prophylaxie de préexposition (PreP): P_{prep}

Ce travail ayant été réalisé avant les résultats définitifs d'IPERGAY, l'efficacité résulte de l'étude IPrEX (121,122) (123).

Nom	**Description**	**Valeurs**
β	Probabilité d'être infecté	*equation 1*
V1	Charge viral médiane	$20\text{-}10^8$
V0	Charge virale initiale	$10^{4.5}$
β0	Probabilité d'infecté avec une charge virale de 4.5 log copies/ml	0,01
n	Nombre d'acte sexuel par pas de temps	1
$P_{preservatif}$	Efficacité du préservatif	toujours: 0.90-0.95 parfois: 0.5-0.80 Jamais:0
P_{prep}	Efficacité de la PreP	0.44

Tableau 2 Variables conditionnant la probabilité de transmission

C Définition des transmetteurs

Tout individu infecté peut être transmetteur, à l'exception des individus au stade SIDA de la maladie qui ne font plus de contact dans le modèle. Nous avons considéré que le diagnostic de ce stade suivait d'un mois la déclaration de la maladie.

D Évolution de la charge viral (CV)

Évolution naturelle de la charge virale

Pendant les trois premiers mois, la charge virale monte vite pour décroitre vers la première valeur de la CV au stade chronique en 3 mois (124–126) (127,128).

Pour la phase chronique de la maladie, il existe une grande variété de données et de publications sur l'évolution naturelle de la charge virale (129,130). Notre modèle est basé sur l'article d'Hubert et al.(131). Ces données françaises sont proches des variations constatées dans d'autres séries (132). Après la première période d'infection (primo-infection), on considère que la charge virale suit une distribution log-normale (4, 0.75). Chaque individu infecté a une charge virale choisie au hasard sur cette loi. Trois groupes d'évolution sauvage de la CV sont définis comme indiqué dans le tableau 3.

Nom	Description	Valeurs	références
VL.wild	Charge viral définie de manière mensuelle	mois 1: $\log(10^7)$ mois 2: $\log(10^6)$ mois 3: (10^5) mois 4: x mois 5 jusqu'à la fin du suivi: • progression standard o mois 6 à 18: x o mois 19 à la fin du suivi: x+(t-18)*0.10/12 • progression lente o mois 6 to 48: x o mois 49 à la fin du suivi: x-(t-48)*0.03/12 • progression rapide: o mois 6 à 18: x o mois 19 à la fin de suivi: x+(t-18)*0.25/12 CV max= $\log(10^8)$	
x	Charge virale initiale après les 3 premiers mois	Tire sur une loi lognormale (4,0.75)	
VL.prep	CV sous PreP	VL.prep = CV.wild-2	(133,134)
VL. treatment	CV sous traitement	Bon répondeur: diminution de 2 log le premier mois puis diminution de 0.5 log mensuel. Répondeur partiel: diminution de 2 log le premier mois puis stabilisation de la charge virale Non répondeur: VL.treatment = VL.wild Cmin= log(20)	(135)

Tableau 3 Variables définissant la valeur de la charge virale de manière mensuelle.

Évolution de la charge virale sous PreP

Il est difficile d'estimer l'influence du PreP sur la CV en cas d'infection survenant sous traitement. Une étude sur les singes suggère que la PreP fait baisser la charge virale de 2 log (136–138). Nous considérons que la PreP est arrêtée après un test de dépistage positif qui est effectué tous les 3 mois dans cette population sous PreP. La PreP est arrêtée si la personne est diagnostiquée séropositive. Le traitement commence le mois suivant.

Évolution de la charge virale sous traitement

Nous avons pris en compte 3 situations selon que la personne prend parfaitement son traitement et ne pose pas de problème thérapeutique (bon répondeur), selon que l'observance et le suivi est partiel (répondeur partiels) ou que le patient ne prenne pas son traitement ou n'est pas suivi (135).

La probabilité d'appartenir à un de ces trois groupes était définie ainsi : : (135)

- bon répondeur: p = 0.83
- Répondeur partiel: p = 0.08
- Non répondeur: p = 0.09

E Évolution des CD4

Nous avons considéré que la CV et le taux de CD4 étaient indépendant. Le taux de CD4 était important pour définir l'évolution vers le stade SIDA. Quatre groupes évolutifs ont été définis avec des valeurs de CD4 initiales différentes et une pente de décroissance également différente (640, 670, 768 et 785 CD4/ml en moyenne au début puis le taux décroit par 249,89,87,73 CD4/mm3 /ans respectivement)

Sous traitement, les CD4 remontent au-dessus de 200/mm2 en un an pour les bon répondeur, Les patients partiellement répondeurs gardaient une valeur constante des CD4.

F Probabilité d'être sous traitement

L'événement classant SIDA fera nécessairement démarrer le traitement. Le traitement est commencé dans le mois qui suit le diagnostic de l'infection par le VIH, conformément aux directives du Traitement à titre préventif (TasP).

G. Évènements classant SIDA sans traitement (table 4)

L'incidence dépend de la charge virale, du taux de CD4 et de l'âge. Elle suit une loi exponentielle (127,128):

$\lambda_{AIDS.1} = 1/4 \exp\{-3.55 + [-0.21*\sqrt{(CD4)}] + 0.71 (\log VL) + 0.024(Age)\}$

$P_{AIDS.1} = [1 - \exp(-1/12 * \lambda_{SIDA.1})]$

La probabilité individuelle de survenue d'évènement classant SIDA suit une loi de Poisson avec l'incidence des événements classant SIDA comme paramètre (132).

H Réalisation des tests de dépistage

Les tests sont réalisés de manière annuelle pour l'ensemble de la population et de manière trimestrielle pour l'ensemble des individus sous PreP. Il est réalisé dans le mois de l'événement classant SIDA.

Name	Description	Value	references
$\lambda_{AIDS.1}$	Paramètre définissant la probabilité d'un évènement classant SIDA à chaque pas de temps.	1⁄4 exp{-3.55 + [-0.21*√(CD4)] + 0.71 (log VL) + 0.024(Age)}	(127,128)
$P_{AIDS.1}$	Probabilité d'un évènement classant SIDA sans traitement	$P_{AIDS.1}=[1-\exp(-1/12*\lambda_{AIDS.1})]$	(127,128)
$\lambda_{AIDS.2}$	Probabilité d'évènement classant SIDA pendant le traitement.	si CD4>250: - If CV<1000: 5/ (739*12) - If CV>1000: 9/ (295*12) Si 100<CD4<=250: 12/ (140*12) SI CD4<=100: 10/ (44*12)	(132)
$P_{AIDS.2}$	Probabilité d'évènement classant SIDA	$P_{AIDS.2}=\text{Exp}(\lambda_{AIDS.2}(e^{t}-1)$	(132)

Tableau 4 Variables définissant la probabilité d'évènement classant SIDA

2.3.2.4 Article publié

HIV transmission and pre-exposure prophylaxis in a high risk MSM population: A simulation study of location-based selection of sexual partners

Olivier Robineau, Annie Velter, Francis Barin, Pierre-Yves Boelle

Objective

In France, indications for pre-exposure prophylaxis (PrEP) for HIV prevention are based on individual-level risk factors for HIV infection. However, the risk of HIV infection may also depend on characteristics of sexual partnerships. Here we study how place-based selection of partners change transmission and the overall efficiency of PrEP.

Methods

We used the PREVAGAY survey of sexual behavior and HIV serostatus in men who have sex with men (MSM) in a Parisian district to look for associations between sexual network characteristics and HIV infection. We then simulated HIV transmission in a high-risk MSM population. We used information about venues visited to meet casual sexual partners (clubs, backrooms or saunas) to define sexual networks. We then simulated HIV transmis- sion in these networks and assessed the impact of PrEP in this population.

Results

In the PREVAGAY study, we found that HIV serostatus changed with the type of venues vis- ited, in addition to other individual risk factors. In simulations, we found similar differences in HIV incidence when the choice of venues visited was not random. The use of PrEP allowed reducing incidence, irrespective of the venues visited by PrEP users. However, with the same amount of PrEP, the number of infections adverted could almost double depending on network structure and venues visited by PrEP users.

Conclusion

This study shows that characteristics of the sexual network structure can strongly impact the effectiveness of PrEP interventions. These should be considered further to refine individ- ual risk assessment and maximize the effect of individual-based prevention policies.

PLOS ONE

RESEARCH ARTICLE

HIV transmission and pre-exposure prophylaxis in a high risk MSM population: A simulation study of location-based selection of sexual partners

Olivier Robineau[1,2,3]*, Annie Velter[4], Francis Barin[5,6], Pierre-Yves Boelle[1]

1 Sorbonne Universités - Univ Paris 06, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, INSERM U1136, Paris, France, **2** Service Universitaire des Maladies Infectieuses et du Voyageur, Centre Hospitalier Gustave Dron, Tourcoing, France, **3** Département des maladies infectieuses, Univ Lille 2, Lille, France, **4** Santé Publique France, Saint Maurice, France, **5** Université François-Rabelais, INSERM UMR966, Tours, France, **6** Centre Hospitalier Régional Universitaire, Centre National de Référence du VIH, Tours, France

* olivier.robineau82@gmail.com

Check for updates

OPEN ACCESS

Citation: Robineau O, Velter A, Barin F, Boelle P-Y (2017) HIV transmission and pre-exposure prophylaxis in a high risk MSM population: A simulation study of location-based selection of sexual partners. PLoS ONE 12(11): e0189002. https://doi.org/10.1371/journal.pone.0189002

Editor: Joan A Caylà, Agencia de Salut Publica de Barcelona, SPAIN

Received: July 30, 2017

Accepted: November 16, 2017

Published: November 30, 2017

Copyright: © 2017 Robineau et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: For the simulation part of this work, all relevant data are within the paper and its Supporting Information files. This simulation work derived from PREVAGAY 2009 study. Variable used to perform logistic regression and correlation analysis between venue visited are included in a Supporting Information files.

Funding: PYB received grant from Agence publique française de recherches sur le sida et les hépatites virales: www.anrs.fr.

Abstract

Objective

In France, indications for pre-exposure prophylaxis (PrEP) for HIV prevention are based on individual-level risk factors for HIV infection. However, the risk of HIV infection may also depend on characteristics of sexual partnerships. Here we study how place-based selection of partners change transmission and the overall efficiency of PrEP.

Methods

We used the PREVAGAY survey of sexual behavior and HIV serostatus in men who have sex with men (MSM) in a Parisian district to look for associations between sexual network characteristics and HIV infection. We then simulated HIV transmission in a high-risk MSM population. We used information about venues visited to meet casual sexual partners (clubs, backrooms or saunas) to define sexual networks. We then simulated HIV transmission in these networks and assessed the impact of PrEP in this population.

Results

In the PREVAGAY study, we found that HIV serostatus changed with the type of venues visited, in addition to other individual risk factors. In simulations, we found similar differences in HIV incidence when the choice of venues visited was not random. The use of PrEP allowed reducing incidence, irrespective of the venues visited by PrEP users. However, with the same amount of PrEP, the number of infections adverted could almost double depending on network structure and venues visited by PrEP users.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

This study shows that characteristics of the sexual network structure can strongly impact the effectiveness of PrEP interventions. These should be considered further to refine individual risk assessment and maximize the effect of individual-based prevention policies.

Introduction

In Western countries, HIV incidence and prevalence remains the largest in men who have sex with men (MSM) [1]. Differences in risk factors for HIV infection [2] contribute to the variability in the reported ranges [3–5]. For example, in a Parisian district where the prevalence of HIV infection was 18% in MSM attending gay venues [6], HIV infection incidence was 4-times higher than the overall incidence in French MSM [7,8]. In addition to differences in individual risk factors, the network of sexual contacts may change the probability of infection by HIV [9,10]. A better understanding of how sexual network structure changes HIV incidence is all the more required as new prevention approaches target all individuals whose personal characteristics put them at high risk. For example, in Western countries, the highly effective pre-exposure prophylaxis (PrEP) [11–13] is prescribed to MSM with a HIV-positive sexual partner *or* recent bacterial STI *or* a high number of sex partners *or* a history of inconsistent condom use *or* in commercial sex workers [14,15]. Individual risk factors are thus already common in this population, making sexual network characteristics increasingly relevant to explain variation in impact.

Place-based social mixing are of importance for communicable disease spread [16]. The type of venues attended by MSM structures the sexual contact network, determining the pool of possible partners and the type of sexual practice [17,18]. Indeed, large variation in the prevalence of HIV has been reported according to venues attended [19–22].

Here, we aimed at evaluating how the venues visited may affect HIV spread and prevention through differences in the network structure. We first examined the correlation between venues visited and HIV serostatus in the PREVAGAY study, including French MSM attending gay venues in a Parisian district. Then, we used simulations to determine the effect of assortativity in venues attended on HIV dissemination. We then studied whether the population effect of PrEP changed according to venues visited by PrEP users.

Material and method

The PREVAGAY survey provided data on HIV serostatus and places visited by MSM (see *Data*). Sexual networks were simulated based on these data (see *Dynamic sexual network modelling*) and under different assumptions for network characteristics that might be of interest in the epidemic spread (see *Allocating partnerships to venues* and *Effect of correlation between venues visited*). The spread of HIV was simulated in these networks (see *HIV progression*, *transmission*, *screening and interventions*). Finally, we compared the effect of PrEP under different scenarios (see *Simulation Scenarios*).

Data

The PREVAGAY survey was conducted in commercial gay venues in Paris in 2009 [6]. MSM were eligible for the survey if they were at least 18 years old, had had sex with men in the previous 12 months, and could read and speak French quite well. Eight hundred eighty-six (886)

MSM reported socio-demographic data and sexual behavior. They categorized the venues they attended as "bar/clubs without backrooms" (clubs), "saunas with sex" (saunas) and "backrooms/darkrooms/videoclubs/sexclubs" (backrooms). For each participant, a blood sample was obtained concomitantly to test for HIV infection.

We used logistic regression to analyze HIV serostatus (positive/negative) according to individual MSM characteristics including number of partners, risky anal intercourse, age and places visited to meet sexual partners. A P-value less than 0.05 was considered as statistically significant. Analyses were conducted with the R software [23]. S1 Appendix is a data file including variables used in this analysis.

Dynamic sexual network modelling

Casual sexual partnerships are short-lived and changing over time. This can be described by "dynamic stochastic networks", where individuals are "nodes" and "sexual partnerships" are edges changing with time. Here, we described sexual partnerships in three layers corresponding to venue types (clubs, saunas, backrooms) and combined those in a final network (Fig 1). We simulated a virtual population of 10000 MSM by sampling with replacement from the original PREVAGAY participants so that the venues visited matched the PREVAGAY data. Then, we used a separable temporal exponential random graph model (ST-ERGM) [24–28] to describe the creation and dissolution of casual sexual partnerships in each venue type. We assumed a rate of dissolution 0.8 day^{-1}, so that 80% partnerships ending in less than 2 days. The rate of new partnerships creation was chosen to balance dissolutions, so that the average number of partnerships in the population remained constant over time. The R package statnet was used to simulate ST-ERGM [29–32].

Allocating partnerships to venues

In the PREVAGAY study, participants reported an overall number of casual partners, but not the place where they met. Therefore, we explored three rules for allocating new partnerships to venues. Rule "EQL" corresponded to allocating partnerships equally between the three venue types; Rule "BKR" corresponded to twice more partnerships in backrooms than elsewhere; and rule "DNS" allocated partnerships with equal density to all venue types (Table 1)–In a network, density is the percentage of actual partnerships (edges) out of the maximum number of potential partnerships.

Effect of correlation between venues visited

We investigated the effect of correlation between venues attended on the cumulated incidence of HIV infection and on the proportion of infection adverted in scenarios with PrEP. To this end, we compared simulations in networks based on the original PREVAGAY data with simulations in networks where each venue type was attended independently. The latter networks, denoted RND, were built by permuting venues attended between individuals at random. Partnerships were allocated using the 3 rules mentioned above, leading to RND/DNS, RND/EQL and RND/BKR simulations.

HIV progression, transmission, screening and interventions

The model for HIV infection history and transmission was similar to previous descriptions [26,33–35]. Without biomedical intervention (PrEP *or* treatment), viral load after infection followed one of three different patterns. Each pattern presented a peak of viral load during the acute phase of infection, then a slow decrease followed by a rise during late stage of infection.

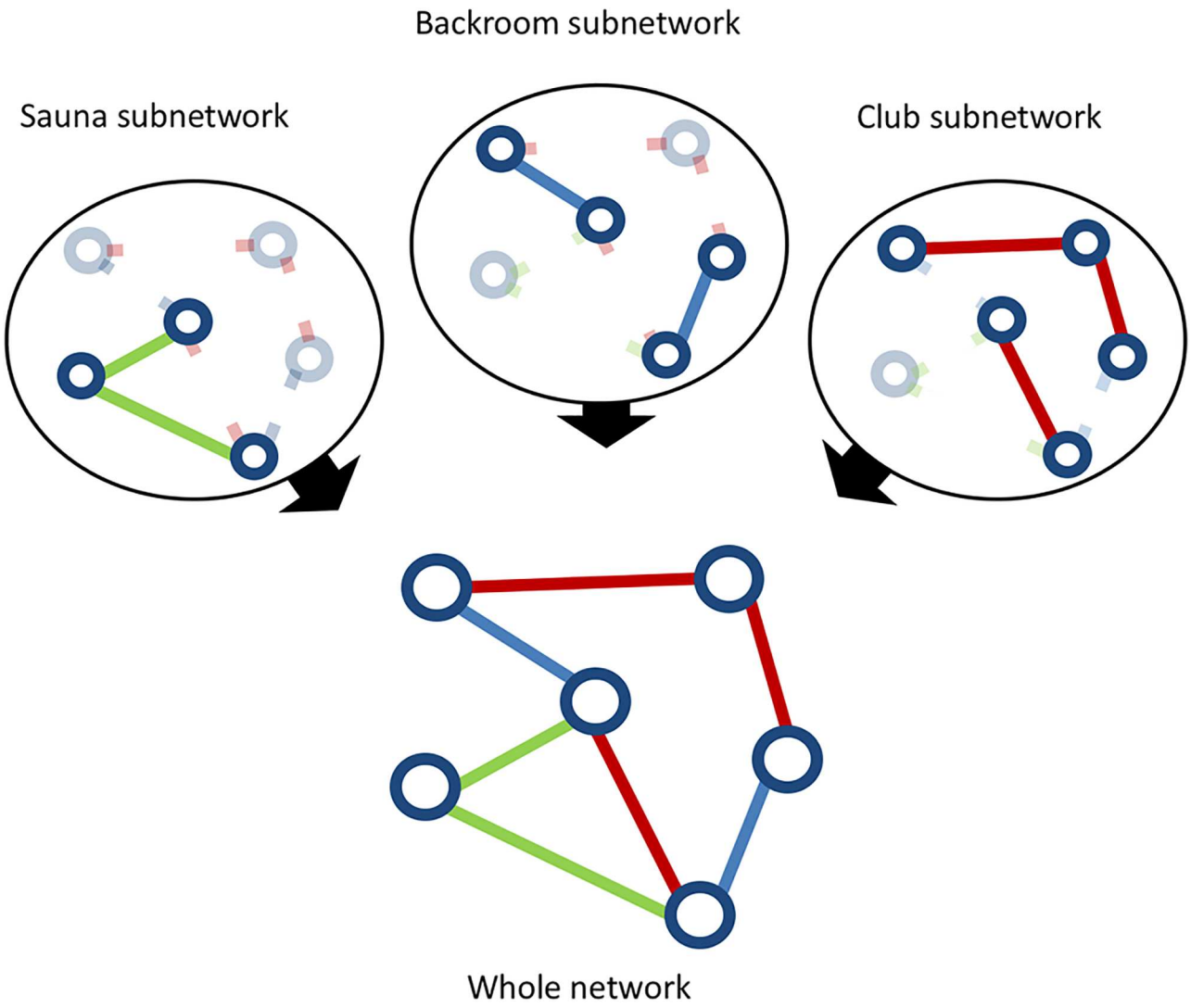

Fig 1. Principle for network construction. Each venue type is modelled using an ST-ERGM as a separate layer (Sauna subnetwork, Backroom subnetwork and Club subnetwork). Subnetworks are overlaid to obtain the whole network.

https://doi.org/10.1371/journal.pone.0189002.g001

In the simulations, viral load was updated monthly according to pre-specified probabilities. With HAART, individuals could be good responder (83% of the cases), partial responder (8%) and non-responder (9%) and the viral load changed accordingly [36]. Transmission occurred between individuals when a HIV infected person was linked to a non-infected person in the sexual network. The probability of transmission was computed as a function of viral load,

Table 1. Rules for allocating partnerships according to venues visited. Density is computed relative to the total number of potential partnerships.

Partnership allocation rule	Description	Proportion of partnerships allocated to each venue type (%) (*Density of contacts, per million potential partnerships*)		
		Club	Backroom	Sauna
DNS	equal density	42% *(0.041)*	16% *(0.041)*	42% *(0.041)*
EQL	equal number	33% *(0.033)*	33% *(0.084)*	33% *(0.033)*
BKR	double in BKR	25% *(0.025)*	50% *(0.126)*	25% *(0.025)*

https://doi.org/10.1371/journal.pone.0189002.t001

modified by interventions (condom use, treatment, PrEP). Quantitative details are given in the S2 Appendix. The baseline scenario for simulation implemented annual screening of all individuals who did not know their HIV status. Highly active anti-retroviral therapy (HAART) was started in the month following a first positive screening test as recommended by international guidelines (treatment as prevention (TasP)) [37]. Condom use was as described in the PREVAGAY study (65%, 23%, 12% of individuals using condom always, inconstantly or never during anal intercourse). In scenario with PrEP, HIV screening frequency was increased to every three months in those receiving PrEP, but no other parameters were modified.

Simulation scenarios

Simulations started with 300 infected individuals chosen at random and ran without PrEP intervention until HIV prevalence reached approximately 18%, similar to the PREVAGAY situation (we verified that the results were little affected by this choice). Then, we introduced PrEP (0%, 10% and 20% of coverage) and investigated PrEP use according to venue types visited. Scenarios were labelled "homogeneous" when PrEP was used irrespective of the venues visited, and "heterogeneous" when PrEP adopters were those visiting a particular venue type. We computed incidence and the proportion of infections adverted over the 36 months following PrEP introduction, relative to the baseline scenario (no PrEP). Finally, we computed the direct effect of PrEP by incidence reduction in individuals belonging to the subpopulation in which PrEP was used (e.g: individuals visiting saunas, clubs or backrooms), and the indirect effect as incidence reduction in other individuals. In all cases, results were averaged over 120 simulations.

Results

Venues visited and HIV serostatus in PREVAGAY

PREVAGAY participants had 17 casual partners/year on average (95% CI 16–19), varying according to venues visited (S1 Table). Participants reported visiting saunas, clubs and backrooms in 73% (646/886), 74% (653/886) and 46% (405/886) of the cases (Fig 2). Visiting backrooms was associated with visiting saunas (OR = 1.93, 95%CI 1.42–2.62) and clubs (OR = 2.88, 95%CI 2.08–3.99). On the contrary, visiting clubs and saunas was less than expected by chance (OR = 0.54, 95%CI 0.38–0.78). The HIV prevalence was 19% (95%CI 16–23) in MSM visiting (at least) clubs and 18% (95%CI 15–21) for those visiting (at least) saunas, and increased to 26% (95%CI 22–30) for those visiting (at least) backrooms ($p<0.001$ for difference with clubs and saunas). Similar, though larger, differences were found in those visiting only one place: prevalence was 12% (95%CI 6–17) for saunas only visitors, 10% (95%CI 5–16) in clubs only visitors and 33% (95%CI 2–64) in backroom only visitors. Backroom visitors were slightly older than in other venues (39.9 years old (95%CI [38.9–40]) vs 37.5 years old (95%CI [36.4–38.5]), $p = 0.002$), although the difference remained quantitatively small. Multivariable analysis showed that older age, higher number of partners, no condom during anal intercourse and visiting backrooms were independently associated with seropositive status (Table 2).

Incidence and prevalence with venues visited in different network settings at baseline

The simulated sexual networks reproduced the overall number of partners observed in the PREVAGAY study. Allocating new partnerships equally between venues (EQL) or preferentially in backrooms (BKR) yielded numbers by venue types close to that in PREVAGAY. (Table 2). The simulated HIV prevalence was also larger in those attending backrooms than

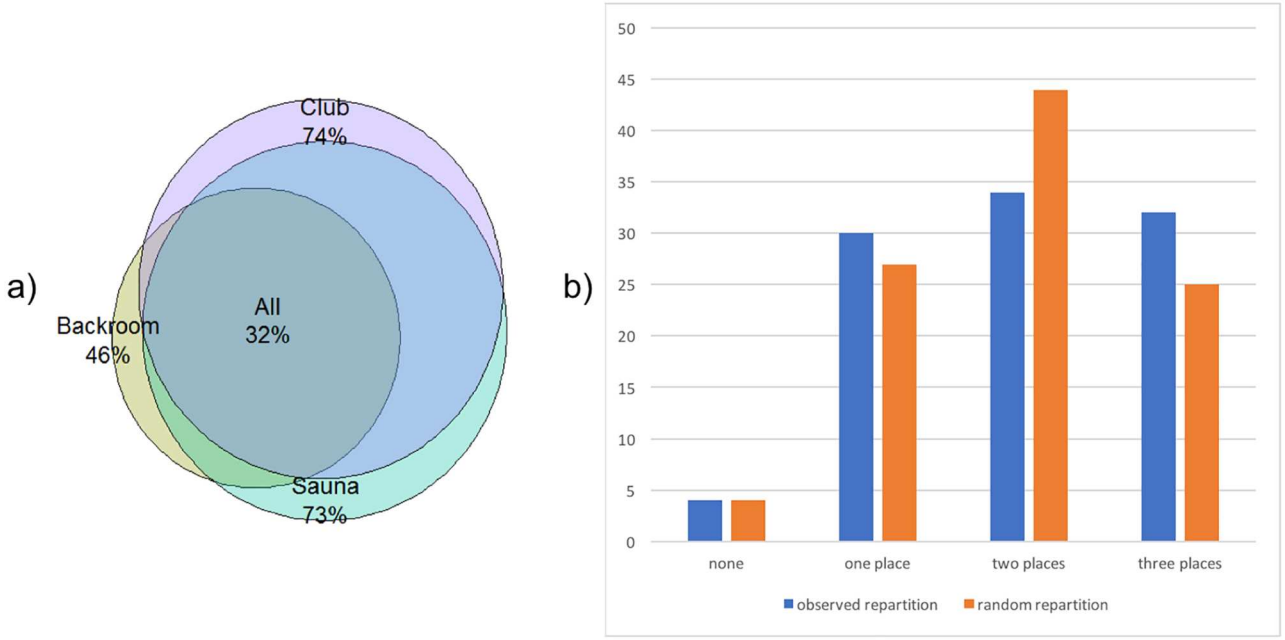

Fig 2. Mixing by venues visited. a) Venue types attended by MSM participating to the PREVAGAY Survey. Percentages are computed with respect to the whole population. b) observed percentage of MSM visiting 0,1, 2 or 3 venue types (PREVAGAY) and expected repartition in the absence of correlation between venues (p<0.001).

https://doi.org/10.1371/journal.pone.0189002.g002

other places. (Fig 3) When the correlation between places visited was removed (RND/* networks), the differences in prevalence were smaller between places. The simulated HIV incidence was between 3 and 4% per year (Fig 4). Allocating more partnerships to backrooms led to an increased circulation of the virus for the same number of contacts: Incidence was 3.08% [2.86–3.32] per year under the DNS allocation and 3.34% [3.05–3.64] per year under BKR allocation, an 8% increase.

The correlation between venues visited led to an increased prevalence by comparison with independent visits to each venues (simulations RND/*). Incidence was 7.5% lower in the absence of correlation between places visited (RND/BKR vs. BKR).

Table 2. Factors associated with positive HIV status in the PREVAGAY study.

Variables	Univariable analysis		Multivariable analysis	
	OR [95% CI]	p	OR [95% CI]	p
Backroom attendance	2.98 [2.07–4.30]	<0.001	2.38 [1.52–3.72]	<0.001
Sauna attendance	0.98 [0.67–1.45]	0.93	-	
Club attendance	1.63 [1.10–2.51]	0.03	-	
Casual partners number (≥ 3vs. <3)	2.70 [1.85–3.93]	<0.001	1.73 [1.10–2.74]	0.02
Condom use during anal sex (No vs. Yes)	3.10 [2.14–4.49]	<0.001	2.87 [1.89–4.37]	<0.001
Age				
≤ 31	1		1	
32–38	2.24 [1.26–3.97]	0.005	1.93 [1.04–3.59]	0.04
38–45	2.49 [1.40–4.40]	0.002	2.06 [1.09–3.87]	0.03
> 45	2.25 [1.27–4.00]	0.005	2.16 [1.15–4.10]	0.02

https://doi.org/10.1371/journal.pone.0189002.t002

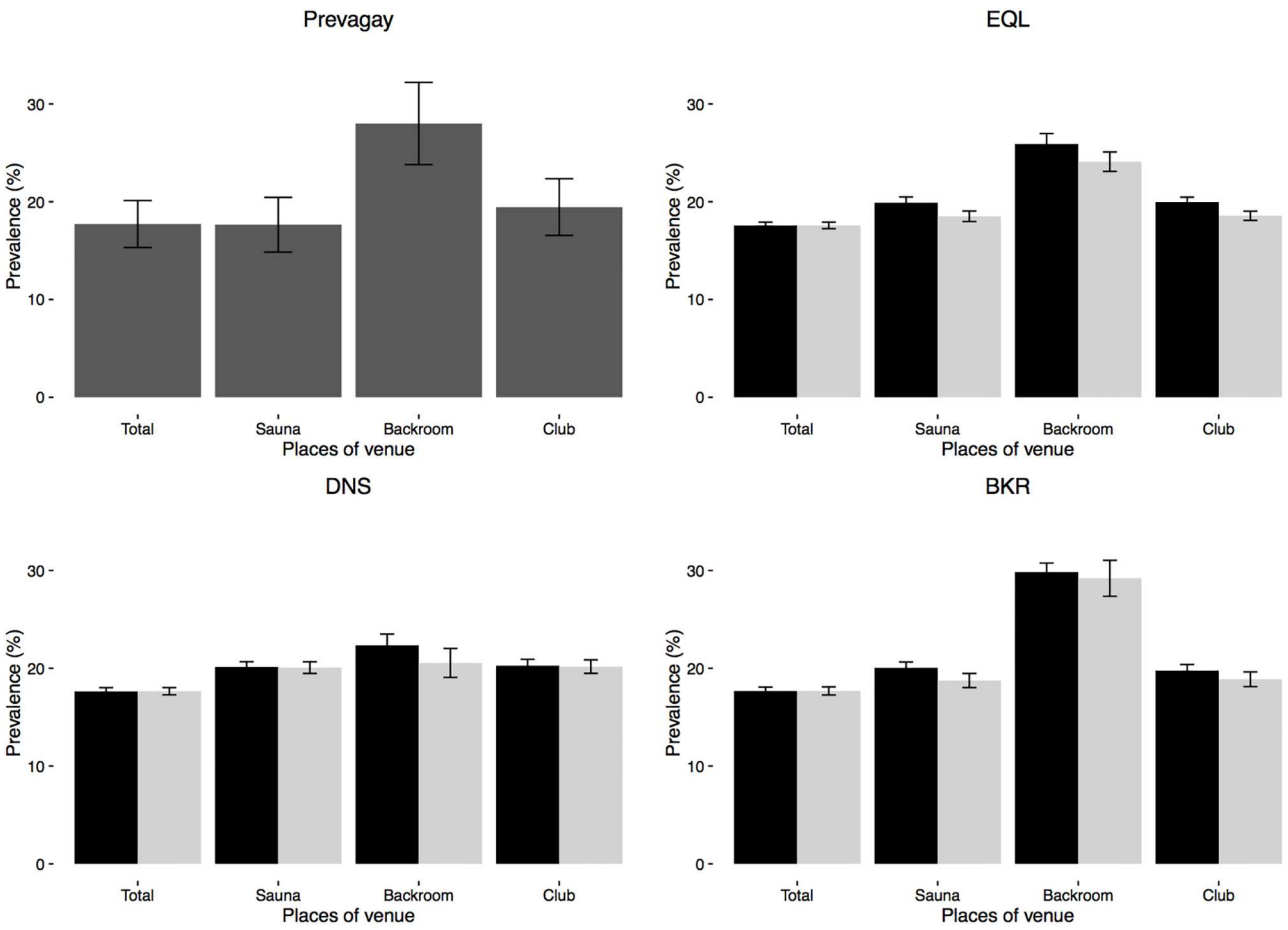

Fig 3. Difference in HIV prevalence in clubs, saunas and backrooms according to assortativity in venues visited. A) PREVAGAY survey. B,C,D) Simulated prevalences for a total of 18% under B) Equal number allocation (EQL); C) equal density allocation (DNS); D) twice more partnerships in Backrooms (BKR). Black bars correspond to the original PREVAGAY data, light gray bars to independence between venues visited (RND/*).

https://doi.org/10.1371/journal.pone.0189002.g003

PrEP effectiveness

Similar results were obtained irrespective of the rules for allocating new partnerships. For the BKR rule, 10% PrEP coverage in MSM used independently of places visited adverted 11.1% (95%CI 10.8–12.5) new infections over 36 months compared to no PrEP. Doubling coverage led to approximately twice more infections adverted over the same period (fraction adverted = 25.4% (95%CI 23.7–27.3)). We found that with the same PrEP coverage, the impact changed according to where the users went: for example, if all the PrEP recipients attended backrooms, 32.2% (95%CI 30.76–33.7) were adverted for an overall coverage of 20% (Table 3). The RND/* scenarios showed that the correlation between places visited reinforced the effect of differential PrEP use according to venues, with 33% more infections adverted in BKR (6.8% = 32.2–25.4, see Table 3) vs RND/BKR(5.1%). Results were similar with the DNS allocation rule.

Direct and indirect effect of PrEP

The proportion of infections adverted was the same in all venue types when PrEP was used independently of places visited (Fig 5A). When users of PrEP went in one place only, the direct

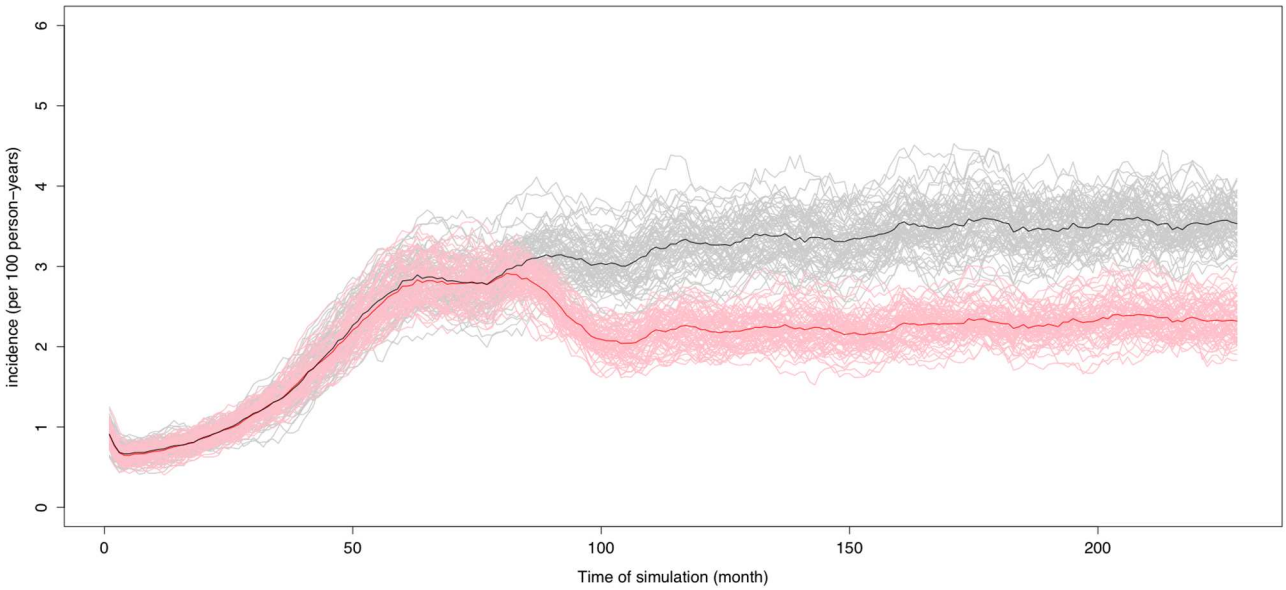

Fig 4. HIV incidence without PrEP (black) and with PrEP (red) in backroom visitors. Results obtained with 20% PrEP coverage in backrooms and the BKR allocation rule.

https://doi.org/10.1371/journal.pone.0189002.g004

effect was the highest when users visited backrooms (see Fig 5B, 5C and 5D). The reduced circulation of HIV in those using PrEP led to indirect protection in non-users, albeit with substantial variation. These indirect effects were the largest when users of PrEP attended backrooms (Fig 5C). In this situation, HIV incidence decreased in those visiting backrooms (from 32.6 to 36.2% reduction) as well as in those not attending backrooms (from 18.3 to 24.4% reduction).

Discussion

We found that differences in the type of venues visited were associated with differences in HIV prevalence, and could modify the efficiency of pre-exposure prophylaxis prevention depending on who uses PrEP.

Table 3. Infection adverted by PrEP after 36 months of intervention. Percentages are relative to the baseline scenario including condom use and TasP in seropositive MSM.

Scenario settings	Networks			
10% coverage	**BKR**	**DNS**	**RND/BKR**	**RND/DNS**
Homogenous	11.1 [10.8–12.5]	11.9 [11.4–12.9]	11.1 [10.8–12.6]	13.1 [11.3–14.1]
Sauna	10.2 [9.8–11.6]	11.2 [1.8–12.6]	11.2 [9.9–12.1]	13.2 [11.4–14.1]
Backroom	15.2 [14.8–16.4]	14.2 [13.8–15.4]	14.6 [13.1–15.9]	13.3 [11.6–14.2]
Club	11.2 [10.6–12.6]	13.2 [12.6–14.1]	11.4 [10.1–12.6]	13.2 [11.6–14.2]
20% coverage				
Homogenous	25.4 [23.7–27.3]	24.0 [22.3–25.7]	25.3 [23.7–27.3]	25.9 [23.6–28.1]
Sauna	25.6 [23.8–27.4]	24.6 [22.9–26.3]	27.0 [24.9–29.0]	26.2 [24.4–28.8]
Backroom	32.2 [30.8–33.7]	28.7 [26.9–30.4]	30.4 [28.2–32.6]	26.6 [24.8–29.3]
Club	25.4 [23.0–27.1]	25.7 [24.1–27.4]	27.5 [25.0–30.0]	27.8 [25.1–30.5]

https://doi.org/10.1371/journal.pone.0189002.t003

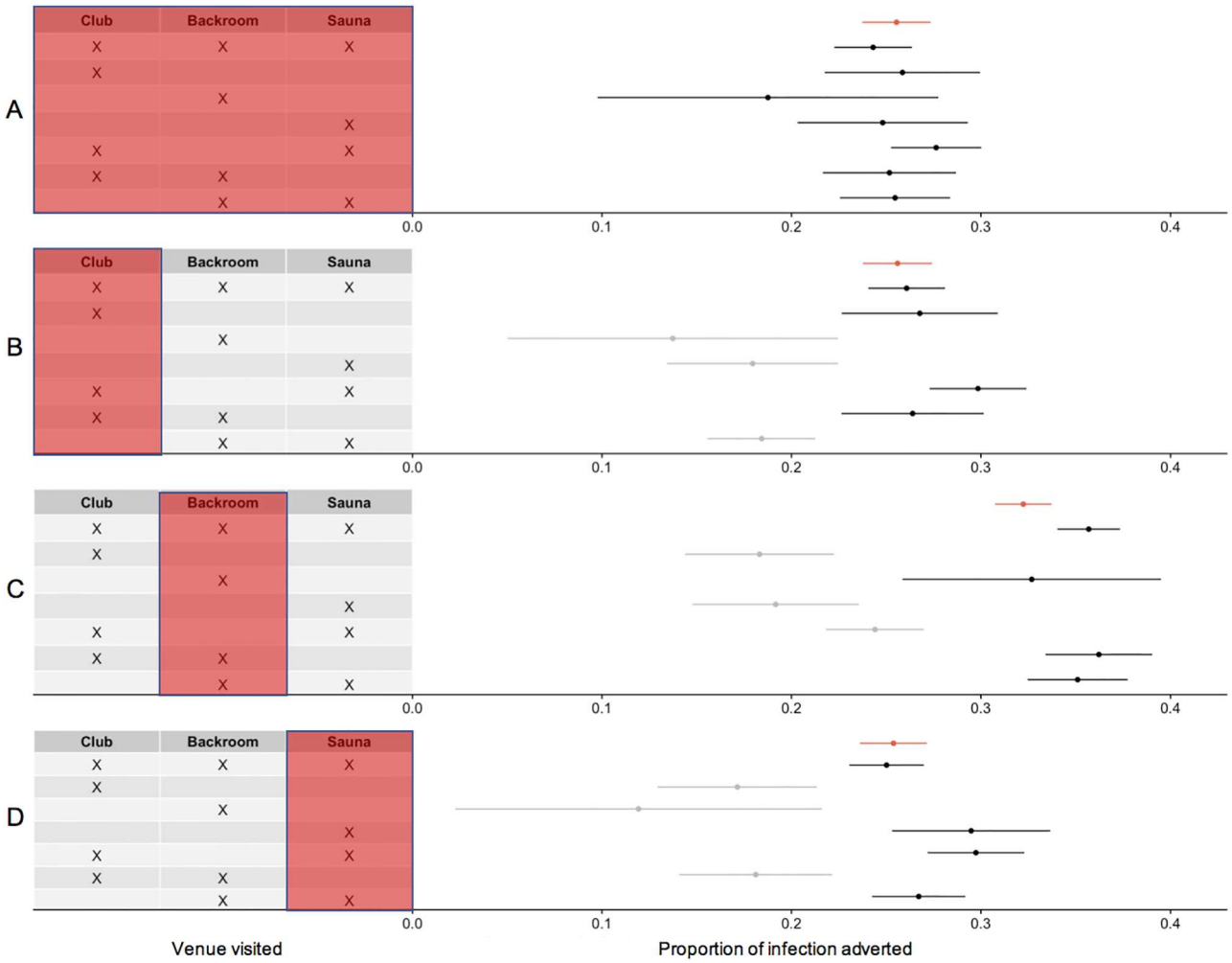

Fig 5. Direct and indirect effect of PrEP coverage: Proportion of infection adverted relative to no PrEP after 36 month. A) homogeneous PrEP use, B) PrEP in saunas, C) PrEP in backrooms, and D) PrEP in clubs. Rows correspond to each pattern of venues visited (Saunas, Backrooms, Clubs), with the proportion of infection adverted. Infection adverted in PrEP users are shown in black (direct effect), and other populations in grey (indirect effect). The average proportion of infection adverted in the whole population is shown in red. Results obtained with 20% PrEP coverage and BKR network.

https://doi.org/10.1371/journal.pone.0189002.g005

HIV transmission and place-based sexual networks

In the PREVAGAY study, the overall HIV seroprevalence (18%) was less than in two venue types (19% in clubs and 26% in backrooms) and equal to prevalence in sauna. This seemingly counter-intuitive result is possible because the venue types attended are not mutually exclusive. Therefore, overall prevalence need not be a weighted average of that in each venue type. More interestingly, prevalence was the highest in backrooms, though backrooms were the least frequently visited venues. This association between backrooms and increased HIV prevalence remained after adjusting for individual characteristics indicative of risky behavior (no condom and $\geq$3 partners). Several mechanisms can contribute to the difference in prevalence between venues. Allocating more partnerships in backrooms led to such differences (BKR vs EQL assumptions). Indeed, it is known that sexual behavior can be riskier in backrooms than in other places [19,38,39]. While finding more partners in one place may contribute to the explanation, we also found that the mere correlation between places attended led to such

differences, as exemplified with the equal density allocation (DNS rule). Indeed, only by removing correlation between places attended could we get no differences between venue types (see DNS/RND situation). Interestingly, changing venues attended over time between MSMs, for example from "backrooms/clubs" to "saunas" or the converse, led to a levelling of prevalence in all venues, irrespective of how partnerships were allocated. This suggest that both number of partners and correlation between places attended are necessary for such difference to occur. We acknowledge that such differences could also arise from reverse causality, if MSM are more likely to attend backrooms once they are seropositive [40]. More studies documenting change in sexual behavior over time, and before and after HIV seroconversion, would be necessary to precise the relative contributions of these mechanisms. Characteristics of sexual networks, including density, homophily (by ethnicity, sexual behavior), concurrency and temporal structure impact the spread of infectious diseases [24,28,41–48]. Recently, assortativity by race or identity group has been explored in MSM as an explanation to differences in HIV spread in selected subpopulation [48,49]. Yet, substantial heterogeneity remains in HIV spread, calling for more research and data on the association of personal and network related risk factor of being infected by HIV.

PrEP prescription and impact

PrEP is a promising opportunity to curb HIV incidence in MSM, and thought to be cost-effective [35,45,50–59]. As of today, international guidelines for PrEP prescription use only individual characteristics known to increase HIV transmission (number of sexual partners, condom use) [14]. In the PREVAGAY study, which describes sexually active members of the gay community in Paris, nearly all participants would qualify for PrEP use according to these standards: all individuals had more than one sexual partner/year and most of them infrequently used condoms. Other individual characteristics (ethnicity, drug use, type of anal intercourse) [60] have been studied to help target those at risk of being infected and who would benefit the most from PrEP. Yet, these models did not investigate the role of the sexual network structure, which is indeed a more recent endeavor in models [26,43,49].

We found that heterogeneous PrEP coverage based on venues visited could outperform homogeneous PrEP coverage. This difference of efficacy between types of coverage depended on partnerships allocation but also on correlation between venues visited. This underlines that network structure, here arising from joint attendance to several venue types, could prove useful to understand differences in reported effectiveness, as well as to better target PrEP recipients. Finally, heterogeneous use of PrEP, especially in backrooms, caused large indirect effects in the MSM population. It is a good example on how individual-based prevention such as PrEP can have an effect on subpopulations who do not benefit of this prevention directly. From a practical point of view, this suggests that places attended could be considered in communication plans to promote the use of PrEP. In this work, we considered 20% coverage for PrEP, when reaching 100% of the target population should be the ultimate goal of public health interventions. Indeed, reaching the population of interest will be key to PrEP efficiency. For example, in the United States of America, there had been 125 000 prescriptions of PrEP 5 years after implementation [61], when the target population was 1 200 000 individuals [62]. In France, 3400 individuals had received PrEP 2 years after implementation, out of an estimated target population of "10 000 to 20 000" individuals [63,64]. Slow adoption may lead to wide variation in coverage, depending on the subpopulations of interest [65]. In this respect, our work may help to shed light on differences in PrEP efficiency in the field, even with the same coverage, but the differences would vanish with 100% coverage

Limitations

We modeled casual sexual partners network according to venue type and on partner allocation in each venue. As such, our model did not account for other individual characteristics that can impact the risk of HIV infection, such as the type of sexual acts or use of prevention, nor on other characteristics affecting new partnerships formation such as age preference or knowledge of serological status. Indeed, individual characteristics will change how partnerships are made within venue types rather than between venues. Additional characteristics could be added to define more precisely the sexual network and the effect of PrEP.

We did not use incidence data to calibrate the model, but only number of partners in the PREVAGAY study and the literature to document the natural history of infection. However, we found that incidence in our simulations were in good agreement with that observed in this population: 3–4% per year in the population (when prevalence reached 18%) vs. 3.8% in the PREVAGAY study [8]. In the same study, 10% of infections were described as "recent", i.e. less than 6 months old, which again compared with the 8%-9% recent infections in the simulations. Interventions were considered once prevalence had reached 18%. We found that the results regarding PrEP and networks were not substantially changed when starting with other prevalence values in a large range between 10 and 30%.

In this work, the venues under consideration did not include social networks, which are now used for finding partners. However, such popular tools may weakly inform on the structure of the sexual network if they are used by a large fraction of MSM. More specific/selective places provide more information on differences in HIV exposure. Another feature of the PREVAGAY study was to focus on French MSM visiting gay venues in Paris. This population is certainly of importance regarding the use of PrEP, given their characteristics, but is not representative of the whole MSM population. Yet, MSM not attending such venues are expected to be at lower risk of infection and spread. Future works needs to be done to build more realistic network implying more mixing variable including other population considered at lower risk, other kind of partnerships and venues.

As mentioned above, PREVAGAY was a cross-sectional study. This made it impossible to tell if MSM attended, say, backrooms before or after seroconversion, and if they changed habits over time. For example condom use is reportedly decreasing which could affect transmission [66]. Longitudinal surveys are needed to document how sexual behavior change over time. For this reason, we only assessed short term outcomes for PrEP efficicency.

Last, real number of sexual contact made in different venues was unknown in our dataset and led us to arbitrary allocation rules. These were chosen with a view to explore different densities of contact between places, providing a sensitivity analysis to the effect of such rules. Allocating the same density to each subnetwork in a network with correlation at random was the only situation where results of heterogeneous and homogeneous coverage were similar. In other more realistic networks settings, we found similar qualitative effect of different PrEP coverage. Homogeneous coverage was less efficient than heterogeneous coverage and assortativity between venues has a role in PrEP epidemiological efficacy.

In conclusion, sexual network structure, here based on venue type preference for finding partners, may alter the community-level efficacy of individual-based prevention methods such as PrEP. As adherence to most prevention methods is difficult to reach and maintain, taking into account network characteristics in the development of prevention campaigns can be of interest. The main objective of future studies is to better define which subpopulation could have the greatest influence in the epidemic process in MSM sexual network, to evaluate if these people have access to PrEP and, if not, how to make this access possible.

Supporting information

S1 Table. Average number of partners and relative HIV prevalence according to venue type visited. The average number of sexual partners by year was computed for the PREVAGAY study and for each rule of partner allocation according to venue type. *Relative to overall prevalence (18%).
(DOCX)

S1 Appendix. Data file containing PREVAGAY variable used in the analysis.
(CSV)

S2 Appendix. Detailed information on the transmission model used in the simulations.
(DOCX)

Author Contributions

Conceptualization: Olivier Robineau, Annie Velter, Pierre-Yves Boelle.

Data curation: Annie Velter, Francis Barin.

Formal analysis: Olivier Robineau, Pierre-Yves Boelle.

Investigation: Olivier Robineau.

Methodology: Olivier Robineau, Pierre-Yves Boelle.

Software: Olivier Robineau.

Validation: Olivier Robineau, Annie Velter, Pierre-Yves Boelle.

Writing – original draft: Olivier Robineau, Pierre-Yves Boelle.

Writing – review & editing: Annie Velter, Francis Barin.

References

1. Jaffe HW, Valdiserri RO, De Cock KM. The reemerging HIV/AIDS epidemic in men who have sex with men. JAMA. 2007; 298: 2412–2414. https://doi.org/10.1001/jama.298.20.2412 PMID: 18042919
2. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet. 2012; 380: 367–377. https://doi.org/10.1016/S0140-6736(12)60821-6 PMID: 22819660
3. Sullivan PS, Rosenberg ES, Sanchez TH, Kelley CF, Luisi N, Cooper HL, et al. Explaining racial disparities in HIV incidence in black and white men who have sex with men in Atlanta, GA: a prospective observational cohort study. Ann Epidemiol. 2015; 25: 445–454. https://doi.org/10.1016/j.annepidem.2015.03.006 PMID: 25911980
4. Goodreau SM, Golden MR. Biological and demographic causes of high HIV and sexually transmitted disease prevalence in men who have sex with men. Sex Transm Infect. 2007; 83: 458–462. https://doi.org/10.1136/sti.2007.025627 PMID: 17855487
5. Sullivan PS, Peterson J, Rosenberg ES, Kelley CF, Cooper H, Vaughan A, et al. Understanding Racial HIV/STI Disparities in Black and White Men Who Have Sex with Men: A Multilevel Approach. PLoS ONE. 2014; 9. https://doi.org/10.1371/journal.pone.0090514 PMID: 24608176
6. Velter A, Barin F, Bouyssou A, Guinard J, Léon L, Le Vu S, et al. HIV Prevalence and Sexual Risk Behaviors Associated with Awareness of HIV Status Among Men Who Have Sex with Men in Paris, France. AIDS Behav. 2013; 17: 1266–1278. https://doi.org/10.1007/s10461-012-0303-1 PMID: 22968398
7. Le Vu S, Le Strat Y, Barin F, Pillonel J, Cazein F, Bousquet V, et al. Population-based HIV-1 incidence in France, 2003–08: a modelling analysis. Lancet Infect Dis. 2010; 10: 682–687. PMID: 20832367
8. Le Vu S, Velter A, Meyer L, Peytavin G, Guinard J, Pillonel J, et al. Biomarker-Based HIV Incidence in a Community Sample of Men Who Have Sex with Men in Paris, France. PLoS ONE. 2012; 7: e39872. https://doi.org/10.1371/journal.pone.0039872 PMID: 22768150

9. Millett GA, Flores SA, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. AIDS. 2007; 21: 2083–2091. https://doi.org/10.1097/QAD.0b013e3282e9a64b PMID: 17885299
10. Smith AMA, Grierson J, Wain D, Pitts M, Pattison P. Associations between the sexual behaviour of men who have sex with men and the structure and composition of their social networks. Sex Transm Infect. 2004; 80: 455–458. https://doi.org/10.1136/sti.2004.010355 PMID: 15572613
11. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med. 2010; 363: 2587–2599. https://doi.org/10.1056/NEJMoa1011205 PMID: 21091279
12. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet Lond Engl. 2015; https://doi.org/10.1016/S0140-6736(15)00056-2 PMID: 26364263
13. Molina J-M, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, et al. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. N Engl J Med. 2015; 373: 2237–2246. https://doi.org/10.1056/NEJMoa1506273 PMID: 26624850
14. Centers for Disease Control and Prevention. Preexposure prophylaxis for the prevention of HIV infection in the United States—2014: a clinical practice guideline, 2014. http://www.cdc.gov.gate2.inist.fr/hiv/pdf/prepguidelines2014.pdf. Accessed 15 February 2016.
15. Prise en charge médicale des personnes vivant avec le VIH. Recommandation du groupe d'experts. Actualisation 2015. Prophylaxie de pré-exposition. https://cns.sante.fr/wp-content/uploads/2015/10/experts-vih_prep2015.pdf.
16. Strömgren M, Holm E, Dahlström Ö, Ekberg J, Eriksson H, Spreco A, et al. Place-based social contact and mixing: a typology of generic meeting places of relevance for infectious disease transmission. Epidemiol Infect. 2017; 1–12. https://doi.org/10.1017/S0950268817001169 PMID: 28625193
17. Binson D, Woods WJ, Pollack L, Paul J, Stall R, Catania JA. Differential HIV risk in bathhouses and public cruising areas. Am J Public Health. 2001; 91: 1482–1486. PMID: 11527785
18. Kerr ZY, Pollack LM, Woods WJ, Blair J, Binson D. Use of multiple sex venues and prevalence of HIV risk behavior: identifying high-risk men who have sex with men. Arch Sex Behav. 2015; 44: 443–451. https://doi.org/10.1007/s10508-014-0352-9 PMID: 25245930
19. Grov C, Hirshfield S, Remien RH, Humberstone M, Chiasson MA. Exploring the venue's role in risky sexual behavior among gay and bisexual men: an event-level analysis from a national online survey in the U.S. Arch Sex Behav. 2013; 42: 291–302. https://doi.org/10.1007/s10508-011-9854-x PMID: 22012413
20. Marcus U, an der Heiden M, Gassowski M, Kruspe M, Drewes J. The impact of meeting locations for men having sex with men on the risk for bacterial sexually transmitted infections: analyses from a cross-sectional online survey. BMJ Open. 2015; 5. https://doi.org/10.1136/bmjopen-2015-009107 PMID: 26537499
21. Grov C, Parsons JT, Bimbi DS. Sexual risk behavior and venues for meeting sex partners: an intercept survey of gay and bisexual men in LA and NYC. AIDS Behav. 2007; 11: 915–926. https://doi.org/10.1007/s10461-006-9199-y PMID: 17206536
22. de Wit JB, de Vroome EM, Sandfort TG, van Griensven GJ. Homosexual encounters in different venues. Int J STD AIDS. 1997; 8: 130–134. https://doi.org/10.1258/0956462971919552 PMID: 9061413
23. R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, AUstria; 2012. http://www.R-project.org
24. Goodreau S, Cassels S, Kasprzyk D, Montaño D, Greek A, Morris M. Concurrent Partnerships, Acute Infection and HIV Epidemic Dynamics Among Young Adults in Zimbabwe. AIDS Behav. 2012; 16: 312–322. https://doi.org/10.1007/s10461-010-9858-x PMID: 21190074
25. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Sanchez J, Grinsztejn B, et al. What Drives the US and Peruvian HIV Epidemics in Men Who Have Sex with Men (MSM)? PLoS ONE. 2012; 7: e50522. https://doi.org/10.1371/journal.pone.0050522 PMID: 23209768
26. Sullivan PS, Carballo-Diéguez A, Coates T, Goodreau SM, McGowan I, Sanders EJ, et al. Successes and challenges of HIV prevention in men who have sex with men. Lancet Lond Engl. 2012; 380: 388–399. https://doi.org/10.1016/S0140-6736(12)60955-6 PMID: 22819659
27. Cassels S, Clark SJ, Morris M. Mathematical models for HIV transmission dynamics: tools for social and behavioral science research. J Acquir Immune Defic Syndr 1999. 2008; 47 Suppl 1: S34–39. https://doi.org/10.1097/QAI.0b013e3181605da3 PMID: 18301132
28. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. Am J Public Health. 2009; 99: 1023–1031. https://doi.org/10.2105/AJPH.2008.147835 PMID: 19372508

29. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. J Stat Softw. 2008; 24: 1548–7660.

30. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. J Stat Softw. 2008; 24: nihpa54860. PMID: 19756229

31. Handcock MS, Krivitsky PN. A Separable Model for Dynamic Networks. Statistics.: 26.

32. Goodreau SM. Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network. Soc Netw. 2007; 29: 231–248. https://doi.org/10.1016/j.socnet.2006.08.001 PMID: 18449326

33. Supervie V, García-Lerma JG, Heneine W, Blower S. HIV, transmitted drug resistance, and the paradox of preexposure prophylaxis. Proc Natl Acad Sci U S A. 2010; 107: 12381–12386. https://doi.org/10.1073/pnas.1006061107 PMID: 20616092

34. Supervie V, Barrett M, Kahn JS, Musuka G, Moeti TL, Busang L, et al. Modeling dynamic interactions between pre-exposure prophylaxis interventions & treatment programs: predicting HIV transmission & resistance. Sci Rep. 2011; 1. https://doi.org/10.1038/srep00185 PMID: 22355700

35. Jenness SM, Goodreau SM, Rosenberg E, Beylerian EN, Hoover KW, Smith DK, et al. Impact of the Centers for Disease Control's HIV Preexposure Prophylaxis Guidelines for Men Who Have Sex With Men in the United States. J Infect Dis. 2016; https://doi.org/10.1093/infdis/jiw223 PMID: 27418048

36. Madge S, Smith CJ, Lampe F, Sabin CA, Youle M, Johnson MA, et al. An audit of viral load in one clinical population to describe features of viraemic patients on antiretroviral therapy. HIV Med. 2008; 9: 208–213. https://doi.org/10.1111/j.1468-1293.2008.00548.x PMID: 18298577

37. Consolidated Guidelines on HIV Prevention, Diagnosis, Treatment and Care for Key Populations– 2016 Update [Internet]. Geneva: World Health Organization; 2016. http://www.ncbi.nlm.nih.gov/books/NBK379694/

38. Rusow JA, Fletcher JB, Reback CJ. Sexual Venue Choice and Sexual Risk-Taking Among Substance-Using Men Who have Sex with Men. AIDS Behav. 2017; 21: 1149–1162. https://doi.org/10.1007/s10461-016-1630-4 PMID: 27905014

39. Grov C, Crow T. Attitudes about and HIV risk related to the "most common place" MSM meet their sex partners: comparing men from bathhouses, bars/clubs, and Craigslist.org. AIDS Educ Prev Off Publ Int Soc AIDS Educ. 2012; 24: 102–116. https://doi.org/10.1521/aeap.2012.24.2.102 PMID: 22468972

40. Grov C, Golub SA, Parsons JT. HIV status differences in venues where highly sexually active gay and bisexual men meet sex partners: results from a pilot study. AIDS Educ Prev Off Publ Int Soc AIDS Educ. 2010; 22: 496–508. https://doi.org/10.1521/aeap.2010.22.6.496 PMID: 21204626

41. Doherty IA, Shiboski S, Ellen JM, Adimora AA, Padian NS. Sexual bridging socially and over time: a simulation model exploring the relative effects of mixing and concurrency on viral sexually transmitted infection transmission. Sex Transm Dis. 2006; 33: 368–373. https://doi.org/10.1097/01.olq.0000194586.66409.7a PMID: 16721330

42. Carnegie NB, Morris M. Size Matters: Concurrency and the Epidemic Potential of HIV in Small Networks. PLoS ONE. 2012; 7. https://doi.org/10.1371/journal.pone.0043048 PMID: 22937011

43. Jenness SM, Goodreau SM, Morris M, Cassels S. Effectiveness of combination packages for HIV-1 prevention in sub-Saharan Africa depends on partnership network structure: a mathematical modelling study. Sex Transm Infect. 2016; https://doi.org/10.1136/sextrans-2015-052476 PMID: 27288415

44. Rocha LEC, Liljeros F, Holme P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. PLoS Comput Biol. 2011; 7: e1001109. https://doi.org/10.1371/journal.pcbi.1001109 PMID: 21445228

45. Jenness SM, Sharma A, Goodreau SM, Rosenberg ES, Weiss KM, Hoover KW, et al. Individual HIV Risk versus Population Impact of Risk Compensation after HIV Preexposure Prophylaxis Initiation among Men Who Have Sex with Men. PLoS ONE. 2017; 12. https://doi.org/10.1371/journal.pone.0169484 PMID: 28060881

46. Jenness SM, Weiss KM, Goodreau SM, Gift T, Chesson H, Hoover KW, et al. Incidence of Gonorrhea and Chlamydia Following HIV Preexposure Prophylaxis among Men Who Have Sex with Men: A Modeling Study. Clin Infect Dis Off Publ Infect Dis Soc Am. 2017; https://doi.org/10.1093/cid/cix439 PMID: 28505240

47. Omori R, Abu-Raddad LJ. Sexual network drivers of HIV and herpes simplex virus type 2 (HSV-2) transmission: a comparative mathematical modeling analysis. AIDS Lond Engl. 2017; https://doi.org/10.1097/QAD.0000000000001542 PMID: 28514276

48. Goodreau SM, Rosenberg ES, Jenness SM, Luisi N, Stansfield SE, Millett GA, et al. Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study. Lancet HIV. 2017; https://doi.org/10.1016/S2352-3018(17)30067-X

49. Mitchell KM, Foss AM, Prudden HJ, Mukandavire Z, Pickles M, Williams JR, et al. Who mixes with whom among men who have sex with men? Implications for modelling the HIV epidemic in southern India. J Theor Biol. 2014; 355: 140–150. https://doi.org/10.1016/j.jtbi.2014.04.005 PMID: 24727187

50. Juusola JL, Brandeau ML, Owens DK, Bendavid E. The Cost-Effectiveness of Preexposure Prophylaxis for HIV Prevention in Men Who Have Sex with Men in the United States. Ann Intern Med. 2012; 156: 541–550. https://doi.org/10.7326/0003-4819-156-8-201204170-00001 PMID: 22508731

51. Koppenhaver RT, Sorensen SW, Farnham PG, Sansom SL. The Cost-Effectiveness of Pre-Exposure Prophylaxis in Men Who Have Sex With Men in the United States: An Epidemic Model. JAIDS J Acquir Immune Defic Syndr. 2011; 58: e51–e52. https://doi.org/10.1097/QAI.0b013e31822b74fe PMID: 21921730

52. Pretorius C, Stover J, Bollinger L, Bacaër N, Williams B. Evaluating the Cost-Effectiveness of Pre-Exposure Prophylaxis (PrEP) and Its Impact on HIV-1 Transmission in South Africa. PLoS ONE. 2010; 5. https://doi.org/10.1371/journal.pone.0013646 PMID: 21079767

53. Gomez GB, Borquez A, Caceres CF, Segura ER, Grant RM, Garnett GP, et al. The Potential Impact of Pre-Exposure Prophylaxis for HIV Prevention among Men Who Have Sex with Men and Transwomen in Lima, Peru: A Mathematical Modelling Study. PLoS Med. 2012; 9. https://doi.org/10.1371/journal.pmed.1001323 PMID: 23055836

54. Paltiel AD, Freedberg KA, Scott CA, Schackman BR, Losina E, Wang B, et al. HIV preexposure prophylaxis in the United States: impact on lifetime infection risk, clinical outcomes, and cost-effectiveness. Clin Infect Dis Off Publ Infect Dis Soc Am. 2009; 48: 806–815. https://doi.org/10.1086/597095 PMID: 19193111

55. Desai K, Sansom SL, Ackers ML, Stewart SR, Hall HI, Hu DJ, et al. Modeling the impact of HIV chemoprophylaxis strategies among men who have sex with men in the United States: HIV infections prevented and cost-effectiveness. AIDS Lond Engl. 2008; 22: 1829–1839. https://doi.org/10.1097/QAD.0b013e32830e00f5 PMID: 18753932

56. Cremin I, McKinnon L, Kimani J, Cherutich P, Gakii G, Muriuki F, et al. PrEP for key populations in combination HIV prevention in Nairobi: a mathematical modelling study. Lancet HIV. https://doi.org/10.1016/S2352-3018(17)30021-8

57. Nichols BE, Boucher CAB, van Dijk JH, Thuma PE, Nouwen JL, Baltussen R, et al. Cost-Effectiveness of Pre-Exposure Prophylaxis (PrEP) in Preventing HIV-1 Infections in Rural Zambia: A Modeling Study. PLoS ONE. 2013; 8. https://doi.org/10.1371/journal.pone.0059549 PMID: 23527217

58. Kasaie P, Pennington J, Shah MS, Berry SA, German D, Flynn CP, et al. The Impact of Pre-Exposure Prophylaxis Among Men Who Have Sex With Men: An Individual-Based Model. J Acquir Immune Defic Syndr 1999. 2017; https://doi.org/10.1097/QAI.0000000000001354 PMID: 28498144

59. Punyacharoensin N, Edmunds WJ, De Angelis D, Delpech V, Hart G, Elford J, et al. Effect of pre-exposure prophylaxis and combination HIV prevention for men who have sex with men in the UK: a mathematical modelling study. Lancet HIV. 2016; 3: e94–e104. https://doi.org/10.1016/S2352-3018(15)00056-9 PMID: 26847231

60. Beymer MR, Weiss RE, Sugar CA, Bourque LB, Gee GC, Morisky DE, et al. Are Centers for Disease Control and Prevention Guidelines for Preexposure Prophylaxis Specific Enough? Formulation of a Personalized HIV Risk Score for Pre-Exposure Prophylaxis Initiation. Sex Transm Dis. 2017; 44: 48–56. PMID: 27898570

61. Gilead Sciences, Inc. Annual Meeting of Stockholders, Westin San Francisco Airport, 1 Old Bayshore Highway, Millbrae, CA. May 10, 2017.

62. Smith DK, Van Handel M, Wolitski RJ, Stryker JE, Hall HI, Prejean J, et al. Vital Signs: Estimated Percentages and Numbers of Adults with Indications for Preexposure Prophylaxis to Prevent HIV Acquisition—United States, 2015. MMWR Morb Mortal Wkly Rep. 2015; 64: 1291–1295. https://doi.org/10.15585/mmwr.mm6446a4 PMID: 26606148

63. Molina J, Pialoux G, Ohayon M, Cotte L, Valin N, Ghosn J, et al. One-year experience with Pre-Exposure Prophylaxis (PrEP) Implementation in France with TDF/FTC JM Molina, G Pialoux, M Ohayon, L Cotte, N Valin5, J Ghosn6, E Cua7, C Pintado1, J Chas2, P Pétour8, G Barriere8, F Durand8, the French RTU Group. IAS. Paris; 2017.

64. COMMISSION DE LA TRANSPARENCE, Avis du 22 mars 2017, https://www.has-sante.fr/portail/upload/docs/evamed/CT-15663_TRUVADA_PIC_INS_AvisPostAud_CT15663.pdf.

65. Weinstein M, Yang OO, Cohen AC. Were we prepared for PrEP? five years of implementation. AIDS Lond Engl. 2017; https://doi.org/10.1097/QAD.0000000000001626 PMID: 28857829

66. Beltzer N, Saboni L, Sauvage C, Lydié N, Semaille C, Warszawski J, et al. An 18-year follow-up of HIV knowledge, risk perception, and practices in young adults. AIDS Lond Engl. 2013; 27: 1011–1019. https://doi.org/10.1097/QAD.0b013e32835e1583 PMID: 23698065

2.3.2.5 Résultats principaux

Ce travail présente trois intérêts.

Premièrement, il montre qu'à partir des réseaux de contacts réalisés à partir de peu de variables, il est possible de reproduire des processus épidémiques semblables aux processus observés dans des données réelles. Ces variables, degrés, lieux de rencontres, partenaires simultanés, peuvent être aisément colligées dans des études égocentriques et leur intégration dans un même modèle le rend déjà complexe.

Deuxièmement, il confirme que les lieux de rencontre ont un intérêt central dans l'étude de l'épidémie du VIH chez les HSH. Ils représentent probablement une « unité de pratique sexuelle». Il est évident que le lien n'est pas causal. Chaque lieu représente des pratiques sexuelles, et des individus peuvent avoir des pratiques différentes selon les partenaires, définis par leur lieu. Le fait que les individus fréquentent plusieurs lieux transforme cette variable en variable de réseau dont l'effet dans la transmission se traduit par des disparités de prévalences.

Troisièmement, il confirme que l'impact à l'échelle populationnelle des mesures de prévention individuelles est dépendant des individus sélectionnés pour en bénéficier, mais aussi du réseau sexuel de ces individus. Selon l'hypothèse que la couverture par la mesure de prévention ne peut être totale, il convient d'intensifier l'incitation à la prévention dans ces populations clefs influençant l'épidémie. De manière intéressante, l'effet réseau persiste même avec des couvertures de près de 60% de la population cible et que, selon les individus sous PreP, l'efficacité peut être doublée en termes de nombre d'infections évitées.

2.3.2.6. Critique du modèle de l'article

Le modèle utilisé est dynamique et représente les liens sexuels journaliers d'une population à hauts risques. Les paramètres utilisés sont extrêmement simples pour la création des différents réseaux secondairement empilés. Introduire des notions d'homophilie par âge, niveau socio-économique et autre serait d'intérêt mais nécessite des informations précises issues de *l'ego* et de ses *alters*. Ces notions ont un intérêt dans les relations longues et/ou avec des individus connus et ne peuvent pas s'appliquer aux partenaires inconnus qui forment une large proportion des contacts dans ce réseau HSH. Le fait que notre modèle trouve dans sa configuration la plus réaliste des prévalences dans les différents lieux de rencontres similaires aux valeurs retrouvées dans l'étude PREVAGAY, suggère qu'un modèle simple permet de décrire les réseaux de contacts sexuels occasionnels, pour peu que les lieux de rencontre soient pris en compte. Ceci est confirmé par des études réalisées au Pérou, aux États-Unis et en Australie où la collecte de donnée de réseau est réalisée de longue date (139–141).

Dans notre modèle, les rencontres réalisées sur internet ne sont pas prises en compte car en 2008, elles n'étaient pas encore majoritaires. Internet est devenu aujourd'hui le premier lieu de rencontre des partenaires sexuels. L'intégration de cette variable est aisée sur le plan computationnel, il ne s'agit que d'un lieu de plus, se superposant aux autres. Cependant, Internet est constitué de nombreux sites de rencontre selon le type de relation recherchée. La prise en compte des lieux de rencontres virtuels nécessitera des informations précises sur les caractéristiques des personnes les visitant.

Le modèle de transmission peut être plus complexe en intégrant d'autres variables modifiant le risque de transmission (circoncision, génétique). L'impact des mesures de prévention est également discutable. L'efficacité de la PreP avait été définie suivant les résultats de l'étude princeps iPrEx (142). Les études postérieures ont démontré une efficacité supérieure de cette mesure allant jusqu'à 95% dans l'essai IPERGAY (143). De plus, la rétention dans le soin n'est

pas modélisée sur des données réelles. Les premiers résultats des cohortes françaises insistent sur le fait que la durée de la prise de la PreP peut être très variable. La question du changement de comportement à l'arrêt de la PrEP se pose également.

L'impact de l'usage d'une mesure de prévention sur la structure du réseau n'est pas non plus pris en compte en l'absence de données. Plusieurs hypothèses semblent pourtant d'intérêt. L'homophilie selon l'usage du préservatif ne fait aucun doute, les personnes utilisant les préservatifs ont des rapports sexuels avec des personnes l'utilisant également. Pour la PreP, l'effet de sa prise sur le choix des partenaires est à développer. L'usage de la PrEP pouvant être considéré comme un indicateur pouvant être déclaré, il est possible d'imaginer que les usagers de la PreP ont préférentiellement des relations sexuelles entre elles. Dans cette situation, l'impact de la PreP à l'échelle populationnelle sera diminué (144).

Plus généralement et sur le plan méthodologique, ces modèles bien que dynamiques dans la dissolution et la création des liens ne présentent pas, pour l'heure, de dynamique de modification comportementale. Plus explicitement, Il est possible de calibrer des modifications comportementales, selon des variables pouvant se modifier au cours du temps, si le nombre de modifications et le nombre d'individus impliqués dans ces modifications sont connus. En revanche, les évènements survenant de manière non prévisible dans leur nombre et/ou leur durée ne peuvent pas être pour l'heure pris en compte comme variables structurant le réseau bien que cela soit théoriquement possible en élaborant de nouveaux paramètre d'ERGM à chaque pas de temps. Outre le problème d'ordre computationnel et donc de temps, ces modifications dynamiques posent le problème de la comparabilité des résultats entre différents scénarios et de la causalité entre nombre d'infections et structure des réseaux, les réseaux ne devenant plus comparables au cours du temps entre deux scénarios différents.

Dans notre exemple de transmission du VIH, il serait possible d'intégrer des modifications de paramètres définissant les liens d'un individu selon des évènements prévisibles et/ou moyennés

comme l'âge ou l'arrêt de la PrEP si ces informations sont connues. Intégrer des modifications comportementales chez les nouvelles PVVIH s'avère plus complexe puisque selon les scénarios, leur nombre est a priori différent et initialement inconnu, l'objectif de ce type de simulation étant de comparer leur nombre selon les différents scénarios.

3. Utilisation des réseaux dans l'estimation d'une prévalence de maladies

3.1. Généralités sur les méthodes d'échantillonnage par les pairs

Les enquêtes épidémiologiques cherchent à obtenir un échantillon d'une population source représentatif de cette même population et idéalement de la population cible. La méthode d'échantillonnage aléatoire est la méthode de référence mais sa mise en pratique en vie réelle est souvent difficile, particulièrement dans les populations dites « difficiles d'accès ». De plus, les individus atteignables facilement ne sont peut-être pas toujours représentatifs de la population cible.

Pour accéder aux populations difficilement atteignables, des outils d'échantillonnage orientés ont été élaborés utilisant les méthodes de recrutement par les pairs. Leur principe est que les individus d'un groupe d'intérêt recrutent d'autres individus du même groupe dans le programme de prévention. Ceci forme des chaines qui suivent les chemins du réseau de contact sous-jacent. Historiquement, ces méthodes ont été développées pour recruter des individus infectés par le VIH dans des parcours de soins mais également des individus à haut risques dans des protocoles de prévention (peer driven intervention) (145–147). Très vite, devant l'efficacité

de ces méthodes à recruter des individus, les équipes à leur origine (Broadhead et Heckathorn) les orientent vers une utilisation en épidémiologie en vue d'estimer la prévalence de pathologie et surtout du VIH dans ces populations difficiles d'accès (22). La méthode générique de recrutement est la méthode « boule de neige » dans laquelle tous les répondants recrutent de nouveaux individus (148). Un cas particulier est l'échantillonnage de recrutement par les pairs ou respondent driven sampling (RDS) qui a l'avantage de bénéficier d'une littérature soutenue quant aux méthodes de redressement des données en vue du calcul d'estimation de prévalence (22).

3.2 Respondent Driven Sampling (RDS)

Les RDS sont un cas particulier d'échantillonnage utilisant les répondants comme recruteurs. Les études débutent par un choix de plusieurs individus, les graines, à partir desquels le recrutement débutera. Il est ensuite proposé à chaque individu recruté de recruter un nombre particulier d'individus (en général 3) qui devront également recruter des pairs, et ce jusqu'à obtention de la taille d'échantillon voulu par vague successive. Le recrutement peut également être maintenu jusqu'à ce que les caractéristiques des nouvelles vagues de répondants soient similaires à la vague précédente. Le recrutement a atteint l'équilibre. Le recrutement s'opère à l'aide de coupons que le recruteur donne aux recrutés, ce qui permet de réaliser le lien entre le recruteur est les recrutés (*ego* et *alters*). La méthodologie de réalisation de ce type d'étude est maintenant bien formalisée et bénéficie d'un groupe de travail méthodologique international validé par l'OMS (149)(150).

Au final, les RDS permettent l'élaboration de jeux de données mixtes. Elles comportent un volet « ego-centrique » avec la description de chaque ego à qui il est demandé de plus en plus

souvent de décrire ses alters. Elles comportent également un volet « sociocentrique » car une partie du réseau est réellement observé.

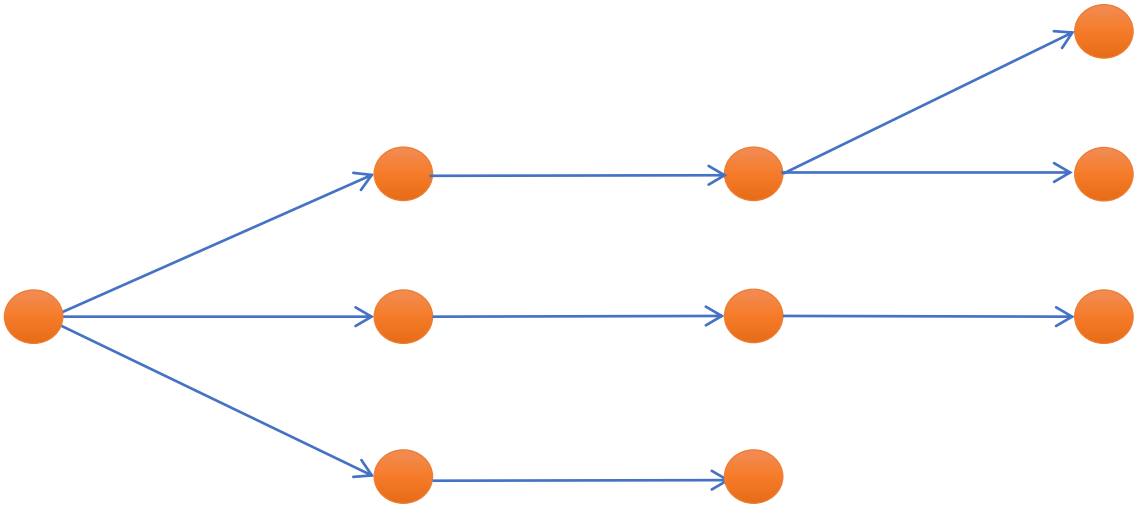

Figure 5 Déroulement d'une étude de type RDS.
Une graine peut recruter jusqu'à N (ici N=3) individus qui pourront recruter autant d'individus jusqu'à obtention de la taille d'échantillon voulue.

La méthode RDS a été développée de manière extensive chez les usagers de drogues intraveineuses (151–153), les HSH (154–158) (159), les transsexuels (160,161) et les travailleurs du sexe (162–164) dans des régions où ces populations sont particulièrement isolées, voire persécutées.

L'élaboration du design de ce type d'étude est complexe car il nécessite de définir le réseau dans lequel l'échantillonnage doit être réalisé. Même avec un objectif d'étude clair comme le calcul de la prévalence du VIH dans une population particulière, la définition des individus à cibler par les recruteurs n'est pas simple. Les enquêtes de prévalence chez les HSH peuvent reposer sur les réseaux de contacts amicaux (recrutez parmi vos amis 3 individus HSH), les réseaux indéfinis sans le lien relationnel précis (recrutez dans votre entourage 3 individus que vous savez HSH) ou les réseaux sexuels (recrutez trois individus avec lesquels vous avez eu

des relations sexuelles au cours des 6 derniers mois). Il est en général proposé d'utiliser des réseaux non définis par le type de relation. Les caractéristiques des relations sont ensuite demandées dans le questionnaire. Le nombre de « relations » du même type est systématiquement demandé dans le questionnaire de recrutement. Cette variable proche du « degré du réseau de contact » est utilisée dans les principales méthodes de redressement des données (cf. infra).

L'autre problématique majeur est l'utilisation de ces données car elles sont biaisées par le recrutement dépendant des individus échantillonnés préalablement. Les calculs de prévalence réalisés sur les données brutes sont donc inexacts, sauf éventuellement si l'échantillon est grand, rendant, entre autre, l'impact des premières vagues dans le calcul de l'estimation négligeable. Cette situation est rare, de si grands échantillons étant réalisables dans des populations faciles d'accès, dans lesquelles la méthode RDS n'a que peu d'intérêt épidémiologique.

Le RDS a été comparé aux autres méthodes d'échantillonnage, que ce soit aux méthodes traditionnelles, basée sur les lieux visités, par méthode boule de neige (165–169). Il résulte de ces travaux en vie réelle ou de simulation que les résultats une fois redressés restent presque toujours superposables à ceux trouvés par les autres méthodes et que lorsque l'échantillon est grand, boule de neige et RDS donnent des résultats similaires. Dans les meilleurs des cas, les RDS semblent permettre d'élargir la distribution des variables collectées, suggérant un échantillonnage diversifié et donc probablement plus représentatif (170). L'autre atout majeur de cette méthode est la rapidité avec laquelle elle permet d'obtenir des échantillons de grande taille.

3.3 Outils de redressement des données de type RDS

3.3.1. Outils historiques

De manière générale, un RDS doit répondre aux caractéristiques suivantes pour qu'il soit le moins biaisé possible :

- Les graines doivent être prises « au hasard » dans la population d'intérêt. Ceci est rendu difficile en vraie vie dans les populations difficiles d'accès. En effet, par définition, la graine doit être « accessible », ce qui ne la rend pas totalement représentative de son réseau.
- Les individus ne doivent pas pouvoir être recrutés plusieurs fois (échantillonnage sans remplacement). Ceci implique que le réseau sous-jacent soit suffisamment grand.
- Le comportement de recrutement est similaire à tous les répondants (définition identique du réseau de recrutement et comprise de la même manière).

L'utilisation des RDS pour estimer une prévalence nécessite ainsi d'utiliser des méthodes de redressement permettant de prendre en compte le biais d'échantillonnage mais aussi de pouvoir prendre en compte les situations fréquentes dans lesquelles les hypothèses sus-citées ne sont que partiellement vraies.

Le premier biais des RDS est donc associé au choix des graines. Il est connu que les premières vagues de recrutement sont très influencées par ce choix (171). La plus simple des méthodes de redressement reste d'éliminer les premières vagues jusqu'au moment où les informations apportées par les nouveaux répondants sont similaires à celles de la vague précédente. Ceci peut n'arriver que très tardivement, comme il sera démontré dans ce travail, entrainant un effort de recrutement important et une perte d'information. Il reste cependant fréquent d'éliminer les

premières vagues (1 à 3), jugées très dépendantes des graines, quelle que soit la méthode de redressement choisie décrite plus bas.

Les méthodes de redressement ont pour objet d'attribuer un poids à chaque individu permettant le calcul de la prévalence. Étant donné que l'échantillonnage est lié au réseau, la plupart des méthodes de redressement utilisent des variables de réseaux disponibles dans le questionnaire proposé aux répondants. Les méthodes les plus utilisées sont celles de Heckathorn ou RDS-1 et celle de Volz-Heckathorn (VH) ou RDS-2. La RDS-1 tient compte des schémas de recrutement entre les sous-groupes et du nombre moyen d'autres membres du groupe cible que les recruteurs connaissent (la "taille du réseau") dans chaque sous-groupe (153,172,173).

La méthode RDS-2 est un estimateur plus récent utilisant l'estimateur généralisé de Horvitz-Thompson ne prenant en compte que le degré individuel déclaré de chaque répondant dans l'estimation de son poids dans le réseau (174). Ce simple estimateur est meilleur que l'estimateur de RDS-1 dans l'estimation de la valeur de point d'une prévalence dans la quasi-totalité des situations où les conditions de réalisation du RDS ne sont pas idéales (sélection des graines, taille de l'échantillon par rapport au réseau)(175).

En effet, l'estimateur de la méthode RDS-2 est asymptotiquement non biaisé pour autant que les hypothèses suivantes soient respectées :

- les répondants déclarent avec précision la taille de leur "réseau" (le nombre d'autres membres du groupe cible qu'ils connaissent) ;
- les répondants recrutent au hasard dans leur réseau
- les répondants ont des relations réciproques avec les membres de la population cible
- le recrutement suit une chaine de Markov et tend vers un équilibre
- l'ensemble des individus d'une population peut être recruté (pas d'individu « isolé »)
- les répondants ne peuvent pas être recrutés deux fois

Ces caractéristiques sont classées dans la table 5 selon leurs liens avec les hypothèses d'échantillonnage nécessaires à la réalisation d'un RDS.

	Hypothèses sur la structure du réseau	**hypothèses sur l'échantillonnage**
Sélection au hasard dans le réseau	Réseau de grande taille ($N >> n$)	*échantillonnage avec remplacement* chaine unique
Indépendance à la graine	*homophilie faible* *peu de goulot d'étranglement* graphes connectés	*nombre de vagues suffisantes*
Comportement de recrutement	*liens réciproques*	*degré mesuré de manière correcte* *recrutement aléatoire*

Tableau 5 Prérequis et hypothèse d'échantillonnage théoriquement nécessaires à l'estimation d'un paramètre à partir de RDS. D'après Gile et col (175).

3.3.2. Méthodes d'estimation utilisant les réseaux

La modèle de VH reste sensible à la sélection des graines, aux variations de recrutement liées à l'homophilie et à la taille de l'échantillon. Il nécessite en plus un nombre important de vague pour être le moins biaisé possible (175). Gile et col ont ainsi développé un nouveau modèle d'estimation rendant possible l'estimation d'une prévalence avant l'obtention de l'équilibre et répondant au moins partiellement aux écueils du modèle VH (176). Ce modèle utilise des réseaux de configurations réalisés à partir du RDS échantillonné. Ces réseaux sont réalisés par des échantillonnages successifs en utilisant une « marche au hasard » dans le RDS. Le poids d'un individu est dépendant de la probabilité qu'il soit échantillonné.

Cet outil est donc plus performant car moins biaisé que les méthodes développées antérieurement. Cependant, il reste sensible à la taille de l'échantillon par rapport à la taille du réseau et aux structures de réseau particulières (homophilie et clustering).

Comme dit plus haut, les RDS peuvent être considérés comme étant une partie observée, ou un échantillon, du réseau de contact formé par la population d'intérêt. Simuler des réseaux de contact pour approcher un réseau à partir duquel le RDS peut avoir a été échantillonné pourrait permettre une estimation des prévalences.

Gile et col ont ainsi développé une méthode d'échantillonnage successif de RDS au sein de réseaux simulés (network-model assisted ou NA). Le principe général est décrit figure 6 (en orange). Les variables nécessaires à l'élaboration de ces réseaux sont les degrés et l'homophilie selon la variable d'intérêt (statut sérologique pour le VIH dans notre cas) (11).

Ces réseaux à partir desquels les RDS sont échantillonnés sont réalisés en utilisant un ERGM prenant comme variables le degré, la distribution du degré selon la variable d'intérêt et l'homophilie selon cette même variable.

Le processus débute par l'estimation des valeurs des statistiques de l'ERGM à partir du RDS observés en utilisant des poids trouvés à partir d'une autre méthode d'estimation (ici SS). Puis :

- M1 réseaux répondant aux statistiques sont simulées
- M2 RDS sont échantillonnés dans les M1 réseaux
- Le poids de chaque individu dans le RDS est calculé selon la fréquence à laquelle des individus de même statut pour la variable d'homophilie et de même degré sont échantillonnés dans les RDS simulés.

Ce processus est répliqué à partir des nouveaux poids jusqu'à obtention de l'équilibre de leur estimation (3 à 5 cycles).

L'estimation de la prévalence est réalisée à partir des derniers poids trouvés.

Les intervalles de confiance sont obtenus en réalisant le même processus N fois (N=100), permettant de décrire la distribution du paramètre.

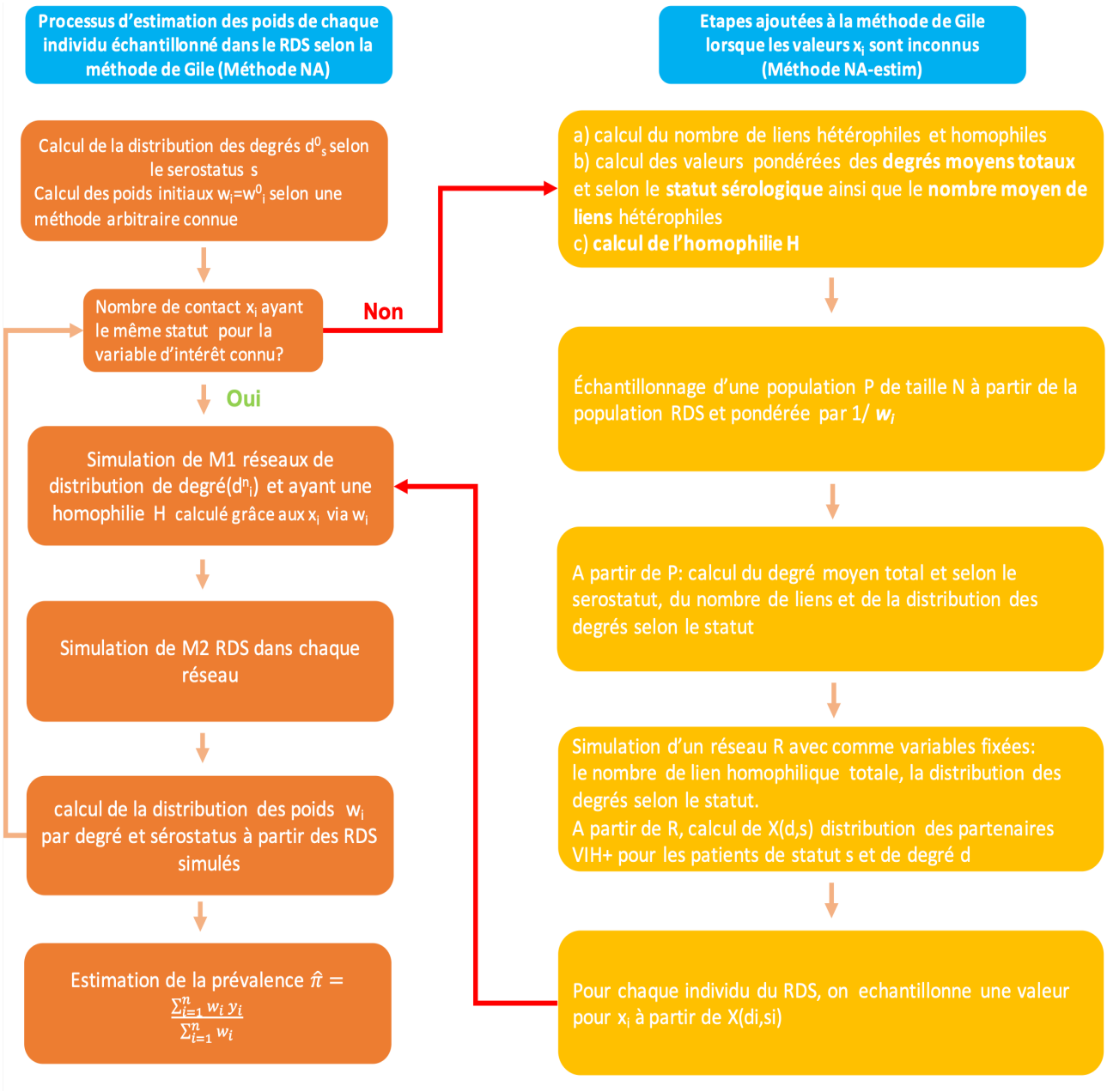

Figure 6 Algorithme permettant l'estimation d'une prévalence selon Gile (en orange). Méthode ajoutée permettant l'estimation d'une homophilie de travail lorsque ce paramètre n'est pas connu (en jaune).

La méthode utilisant les réseaux, comparée aux méthodes précédentes, semble moins sensible aux biais liés au choix des graines, à l'impact de la taille de l'échantillon par rapport au réseau source, à l'impact de variables de réseaux comme une homophilie élevée, la présence de clusters isolés ou en cas de recrutement différentiel (figure 7).

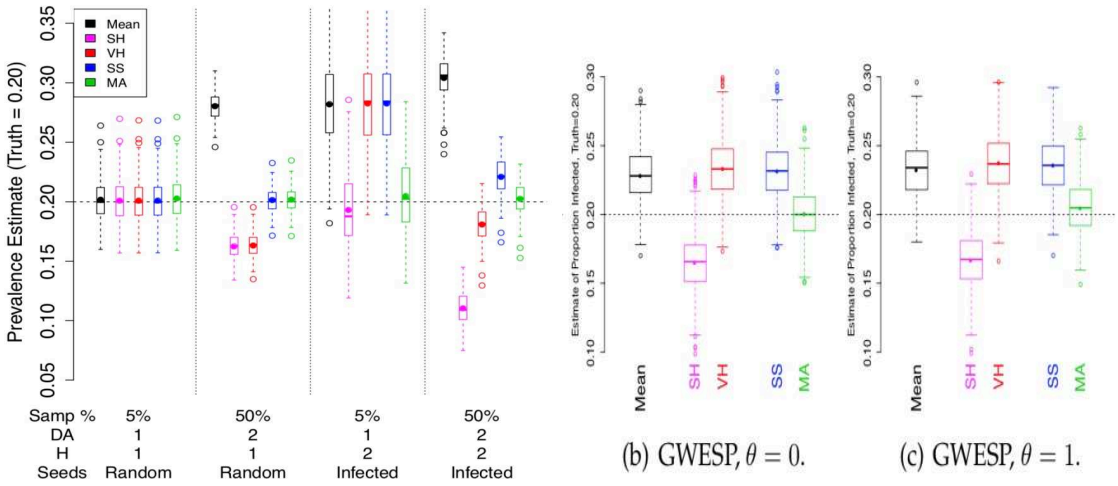

Figure 7 Comparaison des différents estimateurs : selon la taille de l'échantillon par rapport à la taille du réseau (Samp) , le degré d'activité (DA) et les caractéristiques des graines (seeds) et le GWESP défini par la capacité du réseau à réaliser des clusters. Mean : moyenne ; SH : RDS 1 ; VH : Volz Heckathorn ; SS : successive sampling; MA model-assisted. (D'après Gile et col(176))

Comme défini, le modèle NA nécessite que la variable homophilie soit connue. C'est à dire, dans l'exemple du VIH, que le nombre de contacts séronégatifs et séropositifs de chaque *ego* soit connu précisément. Cette variable est parfois disponible dans les études portant sur les usagers de drogues intraveineuses. Elle est en revanche rarement répertoriée lorsque la population cible est définie par ses pratiques sexuelles.

Cette valeur peut être estimée. Gile et col proposent d'utiliser des distributions connues, sous entendue, de la littérature. Ceci est probablement applicable dans les populations d'usagers de drogue intraveineuse car il existe une certaine homogénéité des variables homophiliques dans ces groupes. Cette approche est beaucoup plus discutable dans les réseaux de contact sexuels. La méthode la plus simple est de répartir le degré de contact déclaré selon la distribution de la variable d'intérêt trouvée chez les individus recrutés dans le RDS, selon l'hypothèse que les alters ont été recrutés au hasard dans le réseau de l'égo. Cette estimation peut être affinée en intégrant dans le modèle de Gile de nouvelles étapes (figure 6 en jaune). Cette méthode est l'objet de l'article détaillé plus bas. Brièvement, la distribution des degrés entre les valeurs de

la variable d'intérêt est établie à chaque pas de temps par échantillonnage à partir des réseaux simulés en récupérant les degrés selon le statut de l'ensemble des nœuds du réseau.

La démonstration de l'efficacité de la méthode développée est réalisée à l'aide de simulation de RDS dans des réseaux aux caractéristiques connues (homophilie) et de prévalence de l'infection par le VIH connues. Dans ce travail, A partir de ces RDS, l'estimation de la prévalence de la maladie est réalisée en utilisant l'ensemble des méthodes décrites ci-dessus et la méthode NA modifiée (NA-Estim). Les résultats des différents modèles sont ensuite comparés.

Les modèles ont ensuite été appliqués à un jeu de données réel issu d'une des plus grandes étude RDS réalisées et dont l'objectif était de calculer la prévalence du VIH au sein de la population HSH vivant en zone urbaine au Brésil (159).

3.3 Article

Model-based Respondent- driven sampling analysis for HIV prevalence in brazilian MSM

Olivier Robineau, Marcelo F. C. Gomes, Carl Kendall, Ligia Kerr, André Périssé &

Pierre-Yves Boëlle

Respondent Driven Sampling study (RDS) is a population sampling method developed to study hard-to- reach populations. A sample is obtained by chain-referral recruitment in a network of contacts within the population of interest. Such self-selected samples are not representative of the target population and require weighing observations to reduce estimation bias. Recently, the Network Model-Assisted (NMA) method was described to compute the required weights. The NMA method relies on modeling the underlying contact network in the population where the RDS was conducted, in agreement with directly observable characteristics of the sample such as the number of contacts, but also with more difficult-to-measure characteristics such as homophily or differential characteristics according to the response variable. Here we investigated the use of the NMA method to estimate HIV prevalence from RDS data when information on homophily is limited. We show that an iterative procedure based on the NMA approach allows unbiased estimations even in the case of strong population homophily and differential activity and limits bias in case of preferential recruitment. We applied the methods to determine HIV prevalence in men having sex with men in Brazilian cities and confirmed a high prevalence of HIV in these populations from 3.8% to 22.1%.

SCIENTIFIC REPORTS
nature research

OPEN

Model-based Respondent-driven sampling analysis for HIV prevalence in brazilian MSM

Olivier Robineau[1,2*], Marcelo F. C. Gomes[3], Carl Kendall[4], Ligia Kerr[5], André Périssé[6] & Pierre-Yves Boëlle[1,7]

Respondent Driven Sampling study (RDS) is a population sampling method developed to study hard-to-reach populations. A sample is obtained by chain-referral recruitment in a network of contacts within the population of interest. Such self-selected samples are not representative of the target population and require weighing observations to reduce estimation bias. Recently, the Network Model-Assisted (NMA) method was described to compute the required weights. The NMA method relies on modeling the underlying contact network in the population where the RDS was conducted, in agreement with directly observable characteristics of the sample such as the number of contacts, but also with more difficult-to-measure characteristics such as homophily or differential characteristics according to the response variable. Here we investigated the use of the NMA method to estimate HIV prevalence from RDS data when information on homophily is limited. We show that an iterative procedure based on the NMA approach allows unbiased estimations even in the case of strong population homophily and differential activity and limits bias in case of preferential recruitment. We applied the methods to determine HIV prevalence in men having sex with men in Brazilian cities and confirmed a high prevalence of HIV in these populations from 3.8% to 22.1%.

Respondent-driven sampling (RDS) is a method to sample hard-to-reach populations such as injecting drug users, men who have sex with men (MSM), and sex workers[1]. It uses chain-referral sampling, building on the underlying contact network for recruitment of participants. RDS starts by selecting *seed* individuals from the population of interest. They receive a fixed number of coupons to distribute to individuals in their contact network who meet certain eligibility criteria. In turn, individuals receiving a coupon recruit new participants among their contacts, leading to successive recruitment waves until the target number of individuals for the survey is reached[2]. A drawback of the method is that the final sample is not representative of the target population, introducing bias in naïve estimates of, say, prevalence.

Statistical procedures using generalized Horvitz-Thompson estimators can reduce biases[3,4]. In those, weights are computed according to referral patterns, estimated network size, number of ties between subgroups of interest[5], differences in the number of partners declared by participants[6] or homophily in chain referrals[7]. Weights can also be computed using bootstrap procedures[8]. Yet, numerous issues affect the reliability and validity of RDS estimates because several hypotheses are required[3,9–11]: (1) the population size need to be large compared to the RDS sample; (2) sampling must occur with replacement; (3) population homophily should be weak; (4) seeds should be selected at random.

Recently, the network model-assisted (NMA) method has been shown to increase the robustness in prevalence estimation from RDS data with respect to these assumptions[11]. In this approach, characteristics of the RDS data are used to simulate population networks that resemble the source population. The simulated networks conform to degree distributions and other individual characteristics of participants, but, more importantly, allow

[1]INSERM, Sorbonne Université, Institut Pierre Louis d'Épidémiologie et de Santé Publique, F75012, Paris, France. [2]Service Universitaire des Maladies Infectieuses et du Voyageur, Tourcoing, France. [3]Fundação Oswaldo Cruz (Fiocruz), Programa de Computação Cientifica, Rio de Janeiro, Brazil. [4]Department of Global Community Health and Behavioral Sciences, School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, USA. [5]Department of Community Health, School of Medicine, Federal University of Ceará, Fortaleza, Brazil. [6]Fundação Oswaldo Cruz (Fiocruz), Escola Nacional de Saúde Pública Sergio Arouca (ENSP), Departamento de Ciências Biológicas, Rio de Janeiro, RJ, Brazil. [7]Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP, Hôpital Saint-Antoine, Santé publique, F75012, Paris, France. *email: olivier.robineau82@gmail.com

pair characteristics such as homophily between nodes with similar characteristics, to be taken into account. In simulation studies, the NMA approach fared better than other methods to estimate HIV prevalence in RDS data even in the case of population homophily according to HIV status, when seeds were selected according to HIV status; when the RDS sample size was large relative to the underlying population and the degree distribution changed according to the outcome of interest[11]. The approach however requires knowing the HIV serostatus of all partners of individuals in the RDS sample even those who were not included in the study. This is unlikely to be the case for most RDS studies as serostatus is often determined only in participants. Here, we develop the use of "known patterns" for applying NMA method in this situation[12]. An additional concern is that homophily may not be easily ascribed to the underlying population or to preferential recruitment in RDS data[13]. Assumptions of no preferential recruitment were made in the original method and it is unknown whether accounting for "homophily" of non-specified origin may improve prevalence estimates.

Here, we first show that an iterative approach for applying NMA method leads to significant performance improvement in the case of non-random seed selection in an underlying network with strong homophily and differential activity. We next examine the effect of preferential recruitment on prevalence estimates. We finally compare NMA methods to other methods on RDS data from MSM in Brazil to estimate HIV prevalence.

Material and Methods

We first recall the principle of weighted estimates for estimating HIV prevalence with RDS data and describe an extension to Giles' method for NMA prevalence estimation. Then, a simulation study is presented to investigate the performance of this estimator with three known sources of bias: seed selection, seed dependence, and population homophily. We next investigate the presence of preferential recruitment in the RDS. Finally, we apply the methods to data collected in an RDS study of MSM in Brazil.

Current methods for estimating HIV prevalence in RDS data. An RDS sample for HIV prevalence yields a collection of random variables (y_i) corresponding to the serological status of individuals, with $y_i = 1$ for HIV positive and 0 otherwise. Horvitz-Thompson estimates of the prevalence are computed as $\hat{\pi} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$ where weights w_i correspond to the inverse probability of sampling. Methods differ in the choice of w_i: inverse degree for RDS-I[5], with further modifications in RDS-II[6] and the SS (successive sampling) estimator[8]. The tree bootstrap method allows for another way to compute weights[14]. These approaches are implemented in RDS Analyst, the RDS package for the R Software[15], and the RDS treeboot package[14].

In the NMA method, the computation of weights requires network simulation. This makes it possible and easy to include individual characteristics in modeling contacts, such as infectious status[11], and to account for population homophily, i.e., the differential probability of making links with other individuals based on a specific characteristic. In this approach, population homophily, for example for HIV serostatus, can be defined as the ratio of the observed number of links between discordant HIV serostatus to the number of such links expected by chance[15] where values smaller than 1 correspond to more links than expected between individuals with identical serostatus. This definition corresponds with the parameters used in exponential random graph models (ERGM).

In the original NMA method, it was assumed that the serostatus of all contacts, whether included or not in the RDS, had been reported by each participant. This allowed estimating homophily in the following algorithm for seroprevalence:

Initialization: • compute degree distribution d°_{Y} in RDS data according to serostatus Y • compute initial weights $w_i = w^{\circ}_i$ according to an arbitrary RDS method (here SS method) • compute homophily h° from RDS data
Repeat: Step 1: Simulate M1 networks with degree distribution (d°_i) and homophily h° Step 2: Obtain M2 RDS samples from each simulated network with seeds characteristics as in the original RDS data and with the same number of participants. Step 3: Compute weights distribution w_i by degree and serostatus y based on the simulated RDS averaging over the M1 * M2 RDS samples
Until: • Weights convergence or 5 iterations
Output: • Compute seroprevalence $\hat{\pi} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$ using final weights

Confidence intervals were obtained by the parametric bootstrap using the weights in the last iteration of the procedure a network was simulated using the last weights, RDS was sampled from the network and prevalence estimated from this data[11]. We developed a method to accelerate the Network-Model Assisted method by using configuration network (see supplementary information).

Working with limited information on homophily in the network-assisted method. However, in our experience and the data presented below, participants are unlikely to know or provide information on the serostatus of all of their contacts. Often, accurate information on HIV serostatus about sexual partners is only

City	Sample size	Number of waves	Number of recruited by waves	Number of seeds	Number of infected seeds	Homophily estimate using NMA-Iter
Manaus	848	20	42	10	0	0.55
Recife	351	12	29	10	3	1.01
Salvador	383	20	19	18	2	1.34
Brasilia	344	17	20	10	0	0.46
Campo Grande	351	17	21	7	1	1.71
Belo Horizonte	274	15	18	21	5	0.50
Rio de Janeiro	357	12	30	13	2	1.52
Santos	304	16	19	12	2	1.08
Curitiba	337	13	26	32	2	0.83
Itajai	310	13	24	15	5	0.93

Table 1. characteristics of the RDS in 10 cities of Brazil.

available in those who were recruited in the RDS study. We propose the "NMA-Iter" approach in which imputation of the missing information is first made using the fraction of HIV seropositive observed in direct partners and then updated is updated with other quantities at each iteration. In other words, we start assuming that if 1 out of 3 of the recruited contacts were HIV seropositive for a respondent, then on average 1/3 of all of his contacts were HIV positive, and update this information using the degree distribution computed at each iteration as described in the following algorithm:

Method NMA-Iter

Initialization:

- Initialisation steps of NMA
- For each individual i in the sample set Ii=ki * Pi (infected partners) and Ui=ki* (1-Pi) (uninfected partners) where Pi is the percentage of HIV seropositive recruited partners and ki reported the number of partners.
- Compute Homophily:

$H = 2 \star p_w \star \text{degreew}_{hiv-} \star (1-p_w) * \text{degree}_{whiv+} / (\text{degree}_{woverall} \star \text{degree}_{heterophilic})$

Where p_w is the overall prevalence of HIV infection (weighted mean of x_i), and $\text{degree}_{overall,w}$, $\text{degree}_{hiv-,w}$, $\text{degree}_{hiv+,w}$, and $\text{degree}_{heterophilic,w}$ are, respectively, the mean degree in the overall population (weighted mean of k_i), in those HIV− (weighted mean of k_i when $y_i=0$), HIV+ (weighted mean of k_i when $y_i=1$) and that of heterophilic links (weighted mean of $y_i U_i + (1-y_i) I_i$). All these quantities are post-calibrated with current weights w_i

Repeat:

Step 1–3 of NMA method

Step 4:

- Compute Homophily as above with current weights.
- Simulate a network Nw with degree distribution according to serostatus D(s) and Homophily H.
- From Nw, compute Y(d,s) the distribution of the number of HIV+ partners in individuals with status s and degree d.
- For each patient in the RDS dataset, sample I_i from Y(di, si), compute $U_i = d_i - I_i$

Loop and stop as in NMA method

Simulation study. *Network and RDS simulation.* To investigate the properties of the new proposed estimation procedure (NMA-Iter), we conducted simulation experiments, obtaining RDS data in a population where HIV prevalence was fixed at 20%. We varied the following parameters in the population:

- Population size (Nw). Populations were either 10000 or 1000, so that an RDS sample of size 400 corresponded to either a small (4%) or large (40%) portion of the whole population.
- Differential Activity. DA is defined as the ratio of the average number of partners in HIV seropositive participants and others. It allows simulating differential recruitment according to the characteristic of interest, here HIV serostatus. DA = 1 means no difference, while DA = 2 corresponds to twice as many partners for HIV positive participants.
- Homophily. We finally varied homophily according to the characteristic of interest. Recruitment could be irrespective of serostatus (Homophily = 1) or lead to highly clustered data with twice as many partners of the same status than expected (H = 2).

These characteristics defined six scenarios, for which 150 networks were simulated in each case. Network simulations were performed using the method described in the appendix and exponential random graph models (ERGM) in the R package statnet[16].

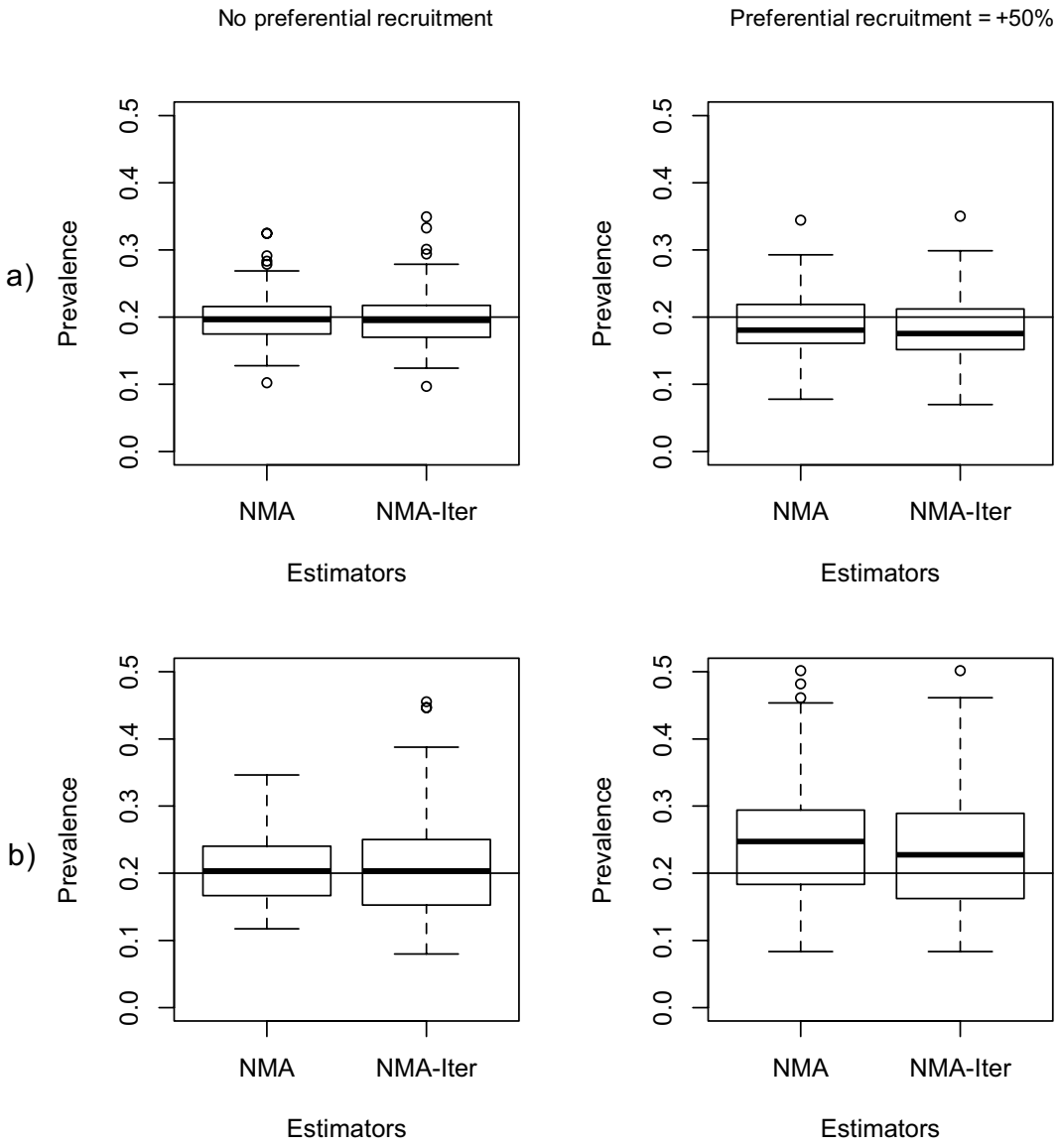

Figure 1. Prevalence estimates with network model assisted method and effect of Preferential recruitment. NMA is the original method requiring full information on serostatus of contacts, NMA-Iter use only serostatus of reported contacts. Preferential recruitment were fixed at no preferential recruitment and +50% of preferential recruitment (left to right). Simulation were performed in networks of size 10000 and: (**a**) population homophily at 1 and DA at 1; (**b**) population homophily at 2 and DA at 2.

RDS samples of size 400 were sampled in each simulated network, using three coupons per participant and an (independent) participation rate of 50%. We started recruitment with seven seeds chosen at random or with seven infected seeds selected, in both cases selected with probability proportional to the degree.

Results were compared graphically by boxplots of the bootstrap distributions. We also computed mean squared errors (MSE) to compare overall errors, including bias and variance of the estimators, for the NMA, NMA-Iter and SS methods.

Effect of preferential recruitment (PR). To investigate the impact of PR in prevalence estimates we obtained RDS samples in the absence of preferential recruitment, then increasing the probability for a partner with the same serostatus to be recruited by 1.5 compared to a partner with discordant serostatus. These situations are described as "No PR" and "+50% PR".

Comparison between methods. We first computed prevalence estimates in the simulated RDS data using the NMA method, assuming full information on homophily. We then computed prevalence estimates using NMA-Iter with limited information on homophily. We used boxplots to show the corresponding distributions.

We then compared estimates from NMA-Iter to estimates obtained using classical RDS analysis methods (RDS-I, RDS-II, SS) in several situations with varying (1) serostatus of seeds (random or infected), (2) RDS

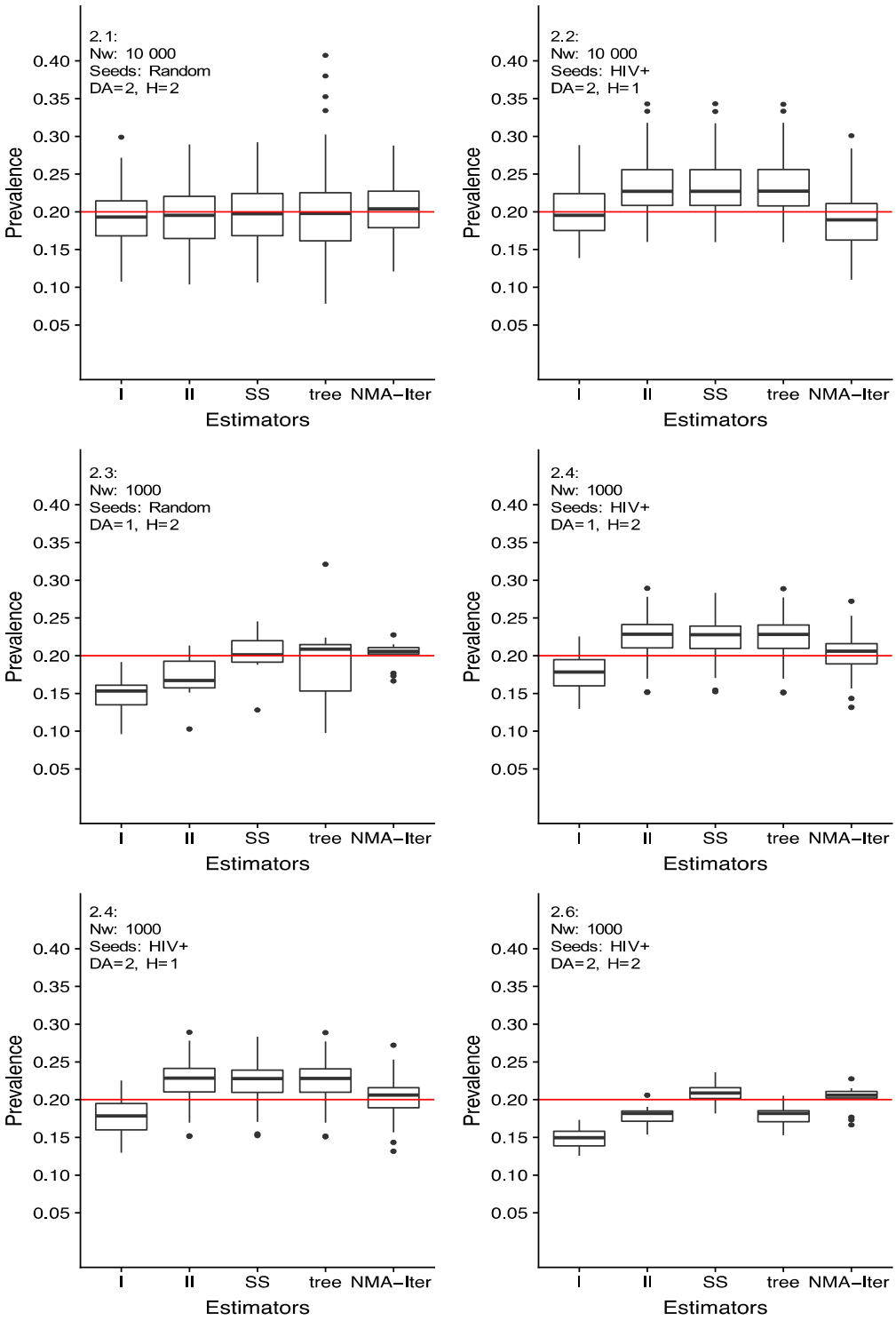

Figure 2. Prevalence estimates using different network patterns in networks of different size. Nw: network size, DA: degree activity, H: Homophily. DA can be at 1 or 2 and H can be at 1 or 2, Nw can be at 10000 or 1000. RDS are of size 400. Preferential recruitment is at 1.

sample size relative to the underlying network: 40% (size 1000) or 4% (size 10000) and (3) magnitude of homophily and DA (1 or 2 for both).

Investigating seed selection. Convergence plots were used to visualize the effect of initial waves on the stability of the final estimates[3]. In this plot, the prevalence estimates are plotted as a function of the accruing number of waves of recruitment. It shows both the effect of initial seed selection and that of sample size.

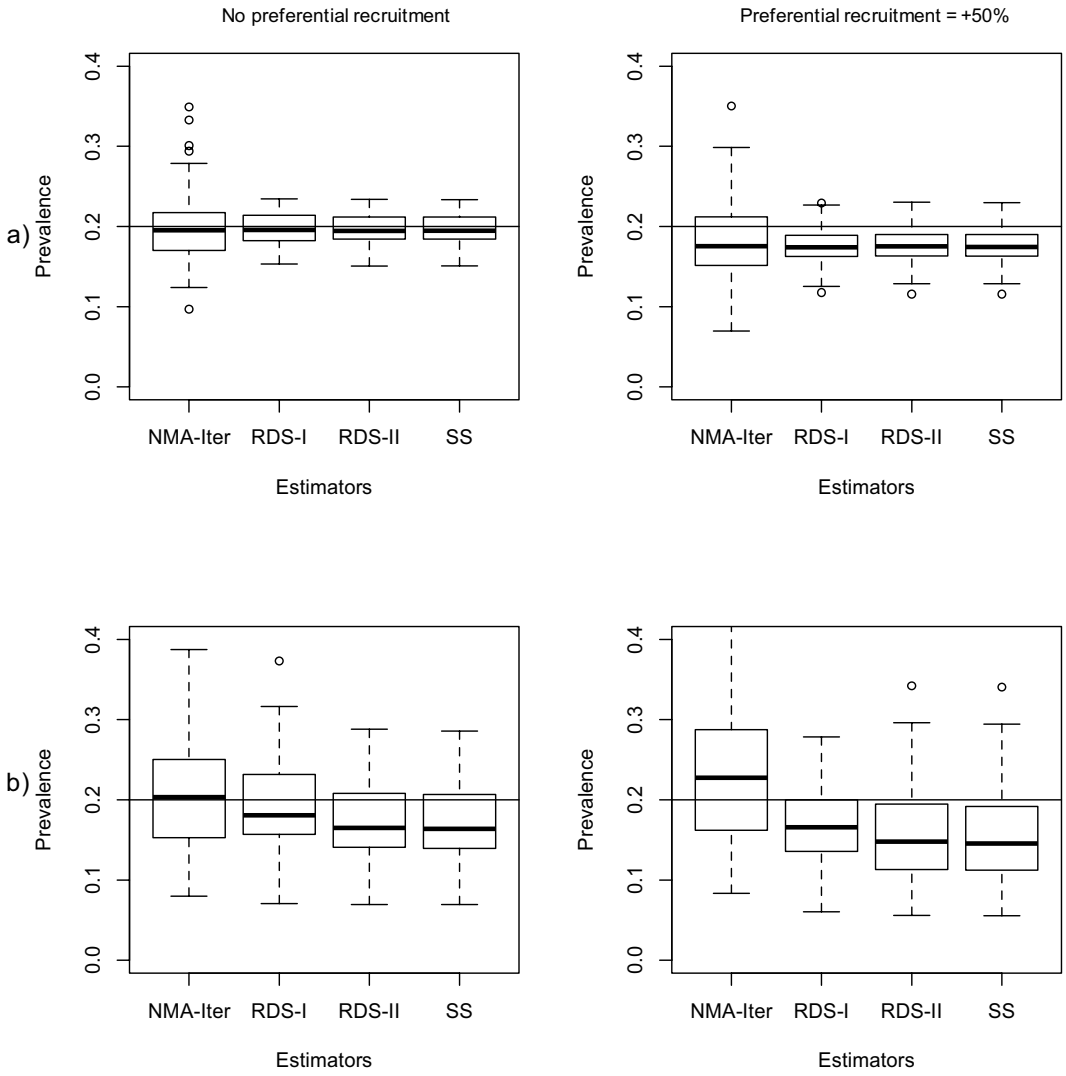

Figure 3. Comparison of different methods to estimates prevalence in RDS with different preferential recruitment. Preferential recruitment were fixed at "No PR" and "+50% PR" (left to right). Simulation were performed in networks of size 10000 and: (**a**) homophily at 1 and DA at 1; (**b**) homophily at 2 and DA at 2.

As it is also customary to drop initial waves in prevalence computations to reduce dependence on the *seeds*, we also show "drop-first convergence plots" where prevalences are computed using all waves but the first ones (i.e., dropping the first, then the first and second, and so on).

We visually inspected convergence plot, and drop-first convergence plots for the SS and NMA-Iter for RDS sampled in networks with no homophily and low DA and in networks with high homophily and high DA. These plots are shown starting at wave 3 and stopping at the last wave of recruitment which might vary from one simulation to another.

Application to HIV prevalence in Brazil. Several RDS studies have been carried out in Brazil in populations with high HIV prevalence[17–23]. Here, we re-analyzed the nationwide survey conducted in 2009 to estimate HIV prevalence in MSM[23]. The survey was conducted in ten Brazilian cities, as described in Table 1.

We applied seven methods to estimate HIV prevalence in each city. For the NMA method, we used a population size of 10000 MSM in each city (a sensitivity analysis showed that estimates did not change significantly above 5000 individuals). We considered that two methods disagreed for a given city when at least one-point estimate was not within the confidence intervals of the other for that same city.

In order to investigate the differences among estimates, we looked at seeds characteristics, population size and examined the convergence plots[3] and drop-first convergence plots.

Ethics. Data from The Brazilian RDS survey was approved by the Brazil National Ethical Research Committee (Comissão Nacional de Ética em Pesquisa, CONEP # 14494). Informed consent was obtained for all participants who signed consent forms for participation to the survey and later use of results[23]. All methods were carried out in accordance with relevant guidelines and regulations[1]. No additional approval was necessary for the current research work.

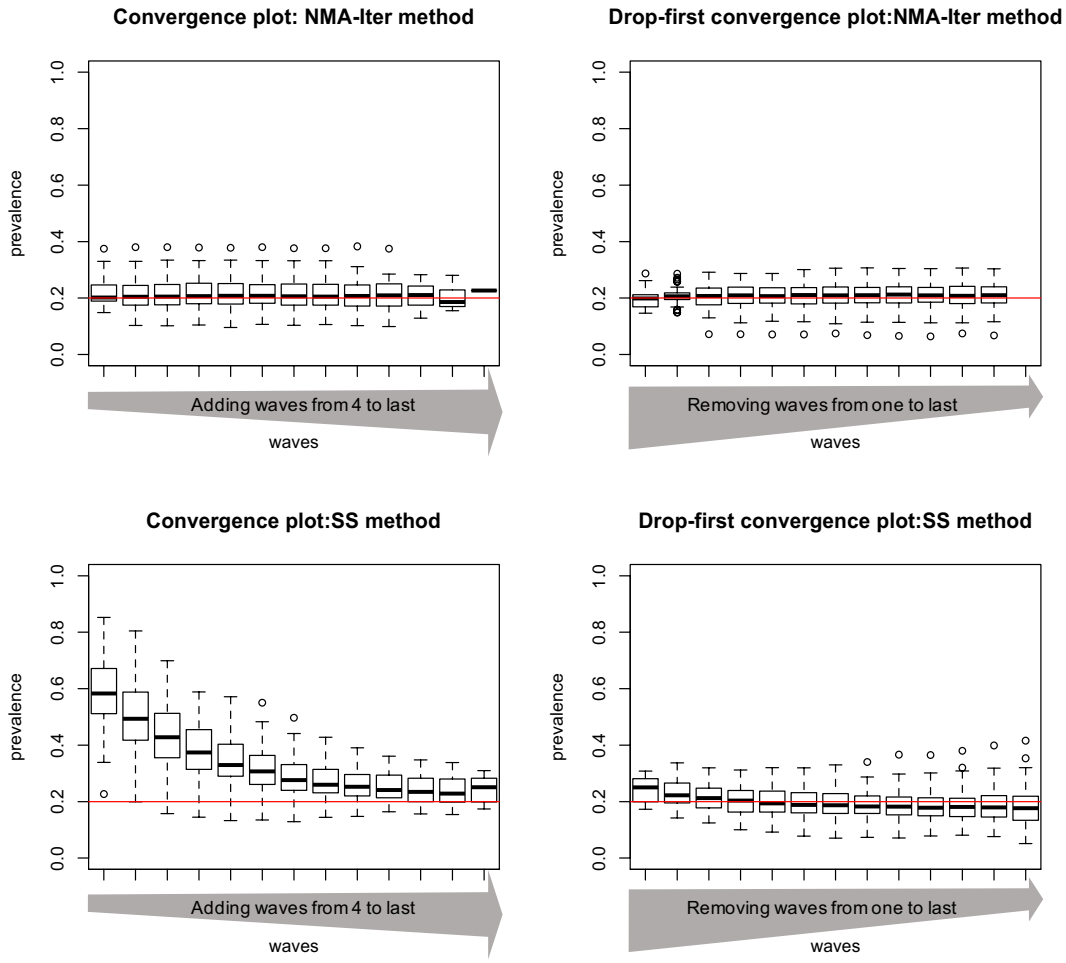

Figure 4. Convergence plot and drop-first convergence plot using SS method (upper part) and NMA-Iter (lower part) in network with high homophily and high degree activity (H = 2 and DA = 2).

City	Previous HIV estimates (logistic regression)	RDS-I	RDS-II	SS	Tree	NMA-Iter
Manaus	8.3 [6–10.9]	8.0 [4.8–11.1]	8.0 [4.9–11.2]	8.0 [5.1–11.0]	6.9 [2.5–12.7]	7.4 [6.5–8.4]
Recife	5.2 [2.7–8.2]	4.5 [1.2–7.8]	5.0 [1.7–8.3]	5.0 [1.6–8.5]	4.5 [1.8–11.4]	5.7 [4.8–8.6]
Salvador	8.9 [5.5–12.7]	6.3 [3.8–8.8]	6.2 [1.8–10.7]	6.3 [1.9–10.6]	5.9 [2.2–13.9]	6.2 [5.5–9.4]
Brasilia	23.7 [16.6–31.5]	13.0 [3.1–22.8]	11.8 [1.8–21.7]	11.8 [4.9–18.7]	19.6 [8.3–34.3]	19.6 [8.0–25.0]
Campo Grande	6.7 [3.5–10.7]	3.3 [1.2–5.5]	3.3 [1.1–5.5]	3.3 [1.2–5.4]	5.0 [1.4–9.4]	4.8 [4.6–5.2]
Belo Horizonte	10.6 [4.2–16.1]	7.4 [0.8–14.0]	7.1 [0.6–13.6]	7.0 [0.8–13.3]	8.9 [3.0–28.0]	10.8 [7.1–12.1]
Rio de Janeiro	18.3 [10.9–24.9]	18.1 [9.6–26.5]	22.2 [13.9–30.4]	22.2 [13.2–31.1]	20.8 [5.0–40.2]	22.1 [17.9–25.5]
Santos	9.0 [5.2–13.5]	2.4 [−0.3–5.0]	3.3 [0.7–5.9]	3.3 [0.3–6.3]	4.0 [1.1–8.1]	3.6 [1.5–5.0]
Curitiba	18.9 [12.6–25.9]	6.3 [2.2–10.5]	7.6 [3.4–11.8]	7.6 [3.3–12.0]	12.8 [7.0–23.2]	9.8 [7.5–15.5]
Itajai	—	11.4 [2.2–20.5]	16.2 [7.0–25.4]	16.2 [6.6–25.8]	15.1 [8.0–25.4]	14.0 [13.4–33.5]

Table 2. Prevalence of HIV in MSM in 10 Brazilian cities based on different RDS analysis methods and from from Kerr and al. previous work.

Results

Simulations. *Performance of the NMA modified method.* The NMA-Iter method allowed estimating prevalence even when serostatus information was limited to RDS participants, with estimates close to the target value (20%; Fig. 1). The results very similar to the original NMA method. The variability of the estimates from the NMA-Iter method did not increase compared to the original NMA, with similar MSE in both cases (1.51×10^{-3} and 1.59×10^{-3} for no homophily nor DA; 2.51×10^{-3} and 2.55×10^{-3} for strong homophily and DA). Additional preferential recruitment led to an increase in bias for both the NMA and the NMA-Iter methods, even though the true prevalence value remained in the confidence intervals for the simulated RDS sample sizes (Fig. 1).

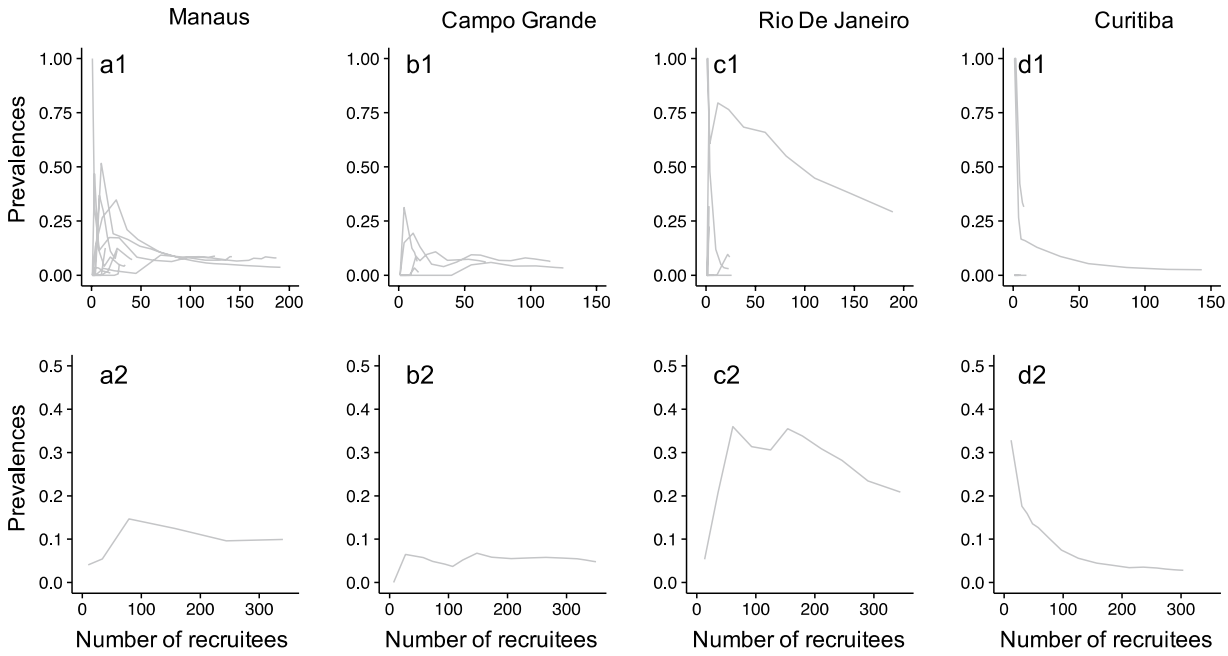

Figure 5. Graphical analysis of convergence. Top row: bottleneck plots showing for each city prevalence estimates evolution per seed during recruitment. Bottom row: convergence plots showing overall prevalence estimate during the recruitment process for each city. All estimates are based on the SS method.

The estimated homophily was close to the nominal in the case of no homophily in the simulated population (H = 1) and no differential activity and no preferential recruitment (estimated value 0.96 [0.87–1.05]). Bias occurred with deviations from the random situation, with downward bias for increasing differential activity (estimated homophily 0.90 [0.81–1.02] for DA = 2 and no PR), and upward bias with preferential recruitment and no DA (1.09 [1.01–1.18] for PR + 50%). With population homophily (H = 2) in the simulated population, this was generally underestimated in the NMA method (estimated homophily = 1.45 [1.29–1.73]), and the estimate increased in case of additional preferential recruitment (1.50 [1.16–2.00] for + 50% PA).

The seroprevalence estimates in the NMA-Iter is shown in Fig. 2 along with other methods. When the population size was large relative to the RDS and seeds were selected at random, all methods yielded similar results even with high homophily and DA. (Fig. 2.1). The bias introduced by selecting only HIV + seeds was the smallest for the RDS-I and NMA-Iter methods (Fig. 2.2). When populations were small (1000 individuals, RDS including 40% of population size), and when seeds were chosen at random, the NMA-Iter method, SS method and tree bootstrap method performed the best under conditions of high homophily and high differential activity (Fig. 2.3). The NMA-Iter remained unbiased when all seed were HIV + in networks with high homophily or DA (Fig. 2.4 and 2.5). Finally, the NMA-Iter method provided the best results for small populations, high homophily and DA, and use of all HIV + seeds (Fig. 2.6).

The presence of preferential recruitment had an impact on seroprevalence estimates for all methods tested (NMA-Iter, RDS-I, RDS-II, SS) (Fig. 3). The bias in the NMA-Iter method remained smaller than in other methods but there was a corresponding increase in variance of the estimates. In the most extreme situation (H = 2, DA = 2, PR = 1.5), the MSE was 12.3×10^{-3} for NMA-Iter and 4.8×10^{-3} for the SS method.

Comparison of SS and NMA-Iter in convergence and drop-first convergence plot. In networks combining differential activity and homophily, the convergence plots showed that the prevalence computed by the SS method was more sensitive to the choice of initial seeds than the NMA-Iter method, with estimates reaching a plateau only after the 10^{th} wave for the SS method and much quicker with the NMA-Iter method (Fig. 4). The drop-first convergence plots even showed that dropping early waves may introduce bias when using the SS method, while the NMA-Iter method remained satisfactory (Fig. 4).

HIV prevalence estimation for Brazilian MSM. Table 2 shows the results for the ten cities using six methods. For most cities, point estimates were between 2 and 10% regardless of the method for estimation, with the exception of Rio de Janeiro and Brasilia where prevalence is around 20%. The results from the different methods did not agree in Recife, Campo Grande, Rio de Janeiro, Santos, Curitiba, and Itajaí (i.e., at least one pair of methods with one-point estimate falling outside the confidence interval of the other method). In Brasilia, variations were substantial for the different methods with correspondingly large CIs. In this city, the point estimate was on the order of 10% for RDS-I, RDS-II, and SS methods, while the Tree bootstrap and NMA-Iter returned a point estimate of around 20%.

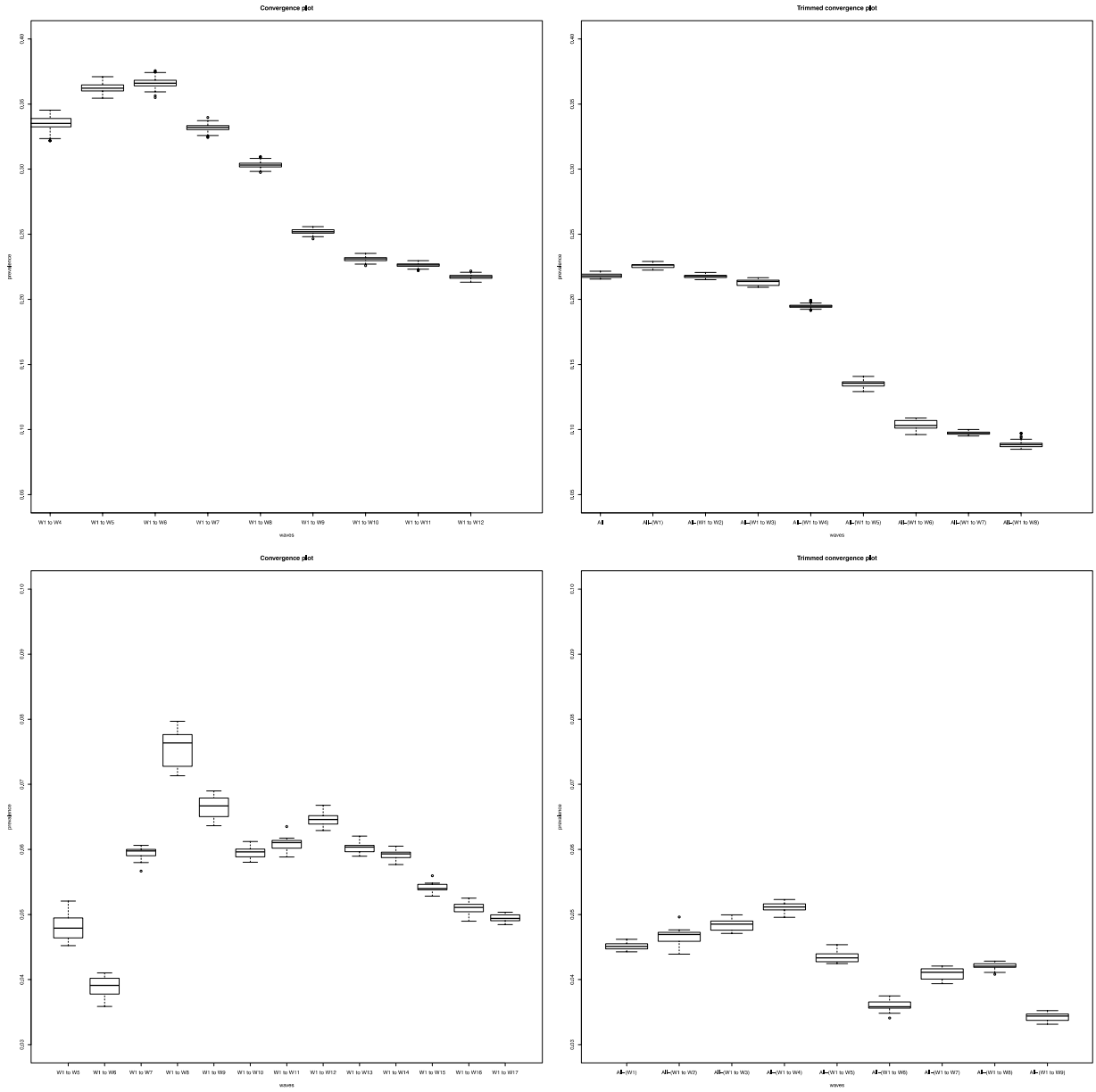

Figure 6. Convergence plot and drop-first convergence plot using NMA-Iter method for the city of Rio de Janeiro (upper part) and Campo Grande (lower part). W: Wave. Results from Rio De Janeiro confirmed the absence of convergence of the estimates. Results from Campo Grande show a convergence around 4% of prevalence. Impact of seed selection is visible up to wave 9.

Seed selection. The proportion of HIV + seeds was high in comparison to the estimated prevalence in Recife (3/10) and Itajaí (5/15) (Table 1). Interestingly, the pattern of results in Itajai was similar to the situation in Fig. 2.4–2.5, with RDS-I yielding smaller prevalence than NMA-Iter and the 3 other methods yielding higher value. For Recife, the results were similar to case 2.6, with NMA-Iter yielding higher estimates than the others.

Population size. In Itajai, it seems that a large part of the MSM population was included in the RDS. Indeed, the RDS included 310 individuals, whereas the MSM population was estimated to range between 700 and 2000 (1 to 3% MSM in adults in a city with 7000 adults)[8]. Once again, the pattern of estimates looked like in Fig. 2 where the RDS fraction was large in the overall population.

Network structure. Homophily calculated using NA-Iter varied from 0.46 for Brasilia to 1.72 in campo grande (Table 1). As seen in the simulation part, non-random homophily might impact results from RDS I and RDS, even if seeds are selected at random. SS and treeboot are biased in the context of seed selection. This is in accordance with our results in Rio de Janeiro, Campo Grande, and Brasilia where homophily values were far from 1.

Convergence of estimates. Figure 5 shows convergence plots for four cities using the SS method. In Manaus, the graphical analysis demonstrated that the sample appeared to achieve convergence (Fig. 5a1,a2). In Campo Grande (Fig. 5b1,b2), convergence plots started with each seed indicated the possibility of distinctive sub-networks being selected, with little indication of convergence. However, in this city, the plots using the NMA-Iter methods suggested otherwise, with little change in the last waves (see Fig. 6).

Lack of convergence of estimates was demonstrated in Brasilia, Santos (not shown) and Rio de Janeiro (Fig. 5c1,c2). In Rio de Janeiro, even at the end of the sample, the slope of the convergence curve appears not to be leveling. It does create doubt that the estimated value is not the final value (Fig. 5c1,c2). This was confirmed for this city by using convergence plot and drop-first convergence plot using NMA-Iter method. Convergence plots showed substantial variations of the estimates before reaching waves 9. Drop-first convergence plot added value by showing that the last waves had still a strong impact on estimates, highlighting that equilibrium was not reached at the end of the RDS process (Fig. 6).

Discussion

In this work, we investigated the use of the Network-Model Assisted method for prevalence estimation using RDS data with limited serostatus information. We found that an iterative method (NMA-Iter) had good properties when information on homophily was limited, even in the case of high differential activity, population homophily and non-random seed selection. An application of these methods to a Brazilian dataset allowed a critical assessment of the reported HIV prevalence among MSM.

NMA-Iter method with respect to other methods. Network modelling is a promising method for estimating prevalence in RDS data as it allows using the characteristics of contact networks, however it requires information on the serostatus of all partners[11,15]. Here, we showed that prevalence estimates obtained when only the subset of individuals included in the RDS sample remained unbiased and did not increase the variability of estimates. This is of importance in practice, as only the serostatus of participants will be available in most RDS samples. An iterative update of all quantities provided better results than a single-step approach (results not shown). The NMA-Iter method kept the advantages already described for NMA compared to other methods in case of seed selection, small network size, population homophily and differential activity. It also compared favorably with the tree bootstrap methods in simulations.

Several assumptions are required to use RDS data for seroprevalence estimation. First, seeds should be selected at random in the population. In practice, one may end up including preferentially, for example, HIV infected individuals as seeds if they can be reached more easily when they seek care or to include a minimum number of cases when seroprevalence is low. The effect of initial seeds should vanish over waves of recruitment in RDS samples[8,10], and this led to recommending dropping early waves before computing estimates. The "drop-first" plot showed that seed selection could affect some methods more heavily than others, was present in the SS method even when the first waves were discarded from the analysis and was much reduced in the NMA-Iter analysis. Using only infected individuals as seeds even led to good performance with the NMA-Iter method irrespective of the characteristics of both the population and the RDS size. The "drop-first" plot is an addition to diagnostic convergence plots[3] for visual inspection of the effect of seed selection.

A second potential limitation is the size of the source population relative to the RDS sample, as the replacement assumption is more likely to be violated when the RDS sample amounts to a large part of this population. In the data we considered, the city of Itajaí may illustrate this situation, as the MSM population size is expected to range between 700 and 2000, while the RDS sample included 310. Simulations for small populations have shown that most methods are biased in this situation, even though it remains a minor contributor to overall bias[4,24]. This was confirmed for the SS estimator and the NMA-Iter that yielded consistent estimates even with sampling fraction as high as 40% of the population.

Last, there are effects of the underlying network structure, for example homophily according to HIV status which has been described among MSM, which affect the RDS data[25,26]. Identifying prevalence estimation methods that can account for it is of importance. The NMA-Iter method, as it estimates homophily from the RDS data and takes it into account for estimation, could reduce bias even with strong homophily and differential activity. These results were however shown with random selection among partners in the RDS data, i.e. no preferential recruitment. The main effect of preferential recruitment according to HIV status should be to increase homophily in the collected data. As both sources of homophily can not be told apart in RDS data, one could assume that estimating a single "working" homophily could preserve the characteristics of the NMA method[13]. In other words, not identifying the cause of homophily may be alright as homophily is a nuisance parameter in the NMA method rather than the target of estimation. Indeed, our simulations showed that the estimated "homophily" parameter in the NMA method changed with both preferential recruitment and population homophily, as it increased with increasing preferential recruitment and population homophily. The simulations showed that the unbiasedness observed with population homophily alone was not preserved in the case of preferential recruitment. NMA-Iter remained less biased than the other methods, but it was also more variable. It would be of interest to examine whether the partial identification reported in Crawford could further improve the NMA results in this respect[13].

HIV prevalence in MSM in Brazil. Stigma and discrimination against MSM remain high in Brazil, especially among the poor and little educated populations[20,27,28]. This makes it difficult to obtain reliable HIV prevalence estimates among MSM and indicates RDS as a choice method for assessing prevalence[17]. Our analysis however showed that estimates changed according to the method and justified looking for those which are the less affected by characteristics of the underlying population. In this respect, we investigated seed selection, population size, recruiting practices, and network structure to find that the NMA-Iter method performed better than other methods. This gives ground to the consideration of its results as more relevant.

Yet, even with the NMA-Iter method, convergence to a well-defined HIV prevalence estimate did not occur in some cities. Too small sample size could be the reason. However, the 3 cities where this was the most apparent (Brasilia, Santos, Rio de Janeiro) did not have particularly small RDS sample size compared with other cities and, furthermore, simulations did not evidence cases of lack of convergence with more than 300 participants. It is more likely that the structure of the population may be at stake. Indeed, recent work indicated that clustering in the population could make prevalence estimates challenging, even after reaching a reasonable sample size[29,30]. As network structure cannot be known *a priori*, it is also reasonable to recommend analyzing RDS data during collection, to detect, as early as possible issues with the recruitment process which could lead to study failure[31].

In this paper we have pursued Gile's injunction to analyze and diagnose sources of bias in RDS studies[3]. The understanding of variation between estimates with different methods is essential to allow for meaningful comparisons. Finding the least biased estimate remains a crucial concern not just for statisticians, but for health authorities to follow the evolution of prevalence over time and to evaluate intervention effectiveness. In Brazil, comparable, national level RDS survey has been conducted among MSM, and been compared to the study reported here[32,33]. Controversy about the high seroprevalences reported, especially among young MSM reiterate the need for improved estimator methods satisfying both methodological and practical concerns to separate methodological and programmatic. How much of the differences found are due to surveillance implementation challenges, estimator methods, changes to the populations and behaviors classified as MSM, or changes in the underlying epidemic? Here we contribute to this discussion, demonstrating that network-assisted model, despite being more time consuming, was the least biased for sources explored in this study.

Limitations. The issues outlined in this work are known shortcomings for the RDS methods[3,4]. Although we have shown that the NMA-Iter method was better than most other methods, we cannot exclude that other unmodelled phenomena contribute to errors in the seroprevalence estimates.

Conclusion

RDS remains at the core of HIV surveillance methods throughout the world for difficult to reach populations. Measuring changes in HIV prevalence is however limited by concerns regarding the variability of findings generated through RDS. The network-assisted model is a powerful method to estimate prevalence using RDS data that may be less biased than other methods. Its use should be encouraged, even if it still requires the use of specific software. These changes are all the more necessary when the effect of treatment as prevention or pre-exposure prophylaxis will have to be quantified in stigmatized populations as MSM in Brazil.

Data availability

Data from the RDS survey are fully available upon request to the author, and are already available online, associated to previous publications.

Received: 4 September 2019; Accepted: 17 January 2020;
Published online: 14 February 2020

References

1. White, R. G. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology for respondent-driven sampling studies: 'STROBE-RDS' statement. *J Clin Epidemiol*, https://doi.org/10.1016/j.jclinepi.2015.04.002 (2015)
2. Heckathorn, D. D. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Soc. Probl.* **44**, 174–199 (1997).
3. Gile, K. J., Johnston, L. G. & Salganik, M. J. Diagnostics for Respondent-driven Sampling. *J. R. Stat. Society. Ser. A,* **178**, 241 (2015).
4. Tomas, A. & Gile, K. J. The effect of differential recruitment, non-response and non-recruitment on estimators for respondent-driven sampling. *Electron. J. Statist.* **5**, 899–934 (2011).
5. Salganik, M. J. & Heckathorn, D. D. Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling. *Sociological Methodol.* **34**, 193–239 (2004).
6. Volz, E. & Heckathorn, D. D. Probability Based Estimation Theory for Respondent Driven Sampling. *J. Off. Stat.* **24**, 79–97 (2008).
7. Heckathorn, D. D. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Soc. Probl.* **49**, 11–34 (2002).
8. Gile, K. J. Improved Inference for Respondent-Driven Sampling Data with Application to HIV Prevalence Estimation. *arXiv:1006.4837 [stat]* (2010).
9. Heckathorn, D. D., Semaan, S., Broadhead, R. S. & Hughes, J. J. Extensions of Respondent-Driven Sampling: A New Approach to the Study of Injection Drug Users Aged 18–25. *AIDS Behav.* **6**, 55–67 (2002).
10. Heckathorn, D. D. Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment. *Sociological Methodol.* **37**, 151–207 (2007).
11. Gile, K. J. & Handcock, M. S. Network Model-Assisted Inference from Respondent-Driven Sampling Data. *J. R. Stat. Soc. Ser. A Stat Soc* **178**, 619–639 (2015).
12. Lu, X. Linked Ego Networks: Improving estimate reliability and validity with respondent-driven sampling. *Soc. Netw.* **35**, 669–685 (2013).
13. Crawford, F. W., Aronow, P. M., Zeng, L. & Li, J. Identification of Homophily and Preferential Recruitment in Respondent-Driven Sampling. *Am. J. Epidemiol.* **187**, 153–160 (2018).
14. Baraff, A. J., McCormick, T. H. & Raftery, A. E. Estimating uncertainty in respondent-driven sampling using a tree bootstrap method. *Proc. Natl. Acad. Sci. USA* **113**, 14668–14673 (2016).
15. Handcock, M. S., Fellows, I. E. & Gile, K. J. RDS: Respondent-Driven Sampling. (2016).
16. Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M. & Morris, M. statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. *J. Stat. Softw.* **24**, 1548–7660 (2008).
17. Kendall, C. *et al.* An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. *AIDS Behav.* **12**, S97–104 (2008).
18. Szwarcwald, C. L., de Souza Júnior, P. R. B., Damacena, G. N., Junior, A. B. & Kendall, C. Analysis of data collected by RDS among sex workers in 10 Brazilian cities, 2009: estimation of the prevalence of HIV, variance, and design effect. *J. Acquir. Immune Defic. Syndr.* **57**(Suppl 3), S129–135 (2011).

19. Toledo, L. *et al*. Putting respondent-driven sampling on the map: insights from Rio de Janeiro, Brazil. *J. Acquir. Immune Defic. Syndr.* **57**(Suppl 3), S136–143 (2011).
20. Sabidó, M. *et al*. Sexual Violence Against Men Who Have Sex with Men in Brazil: A Respondent-Driven Sampling Survey. *AIDS Behav.* **19**, 1630–1641 (2015).
21. Pinheiro Júnior, F. M. L. *et al*. Risk factors associated with resistance to HIV testing among transwomen in Brazil. *AIDS Care* **28**, 92–97 (2016).
22. Grinsztejn, B. *et al*. Unveiling of HIV dynamics among transgender women: a respondent-driven sampling study in Rio de Janeiro, Brazil. *Lancet HIV.* **4**, E169–E176 (2017).
23. Kerr, L. R. F. S. *et al*. HIV among MSM in a large middle-income country. *AIDS* **27**, 427–435 (2013).
24. Barash, V. D., Cameron, C. J., Spiller, M. W. & Heckathorn, D. D. Respondent-Driven Sampling–Testing Assumptions: Sampling with Replacement. *J. Off. Stat.* **32**, 29–73 (2016).
25. Amirkhanian, Y. A. Social networks, sexual networks and HIV risk in men who have sex with men. *Curr. HIV/AIDS Rep.* **11**, 81–92 (2014).
26. Klovdahl, A. S. *et al*. Social networks and infectious disease: the Colorado Springs Study. *Soc. Sci. Med.* **38**, 79–88 (1994).
27. Magno, L. *et al*. Factors associated with self-reported discrimination against men who have sex with men in Brazil. *Rev. Saude Publica* **51**, 102 (2017).
28. Chard, A. N., Finneran, C., Sullivan, P. S. & Stephenson, R. Experiences of homophobia among gay and bisexual men: results from a cross-sectional study in seven countries. *Cult. Health Sex.* **17**, 1174–1189 (2015).
29. Rocha, L. E. C., Thorson, A. E., Lambiotte, R. & Liljeros, F. Respondent-driven sampling bias induced by clustering and community structure in social networks. *arXiv:1503.05826 [physics, stat]* (2015).
30. Wejnert, C., Pham, H., Krishna, N., Le, B. & DiNenno, E. Estimating design effect and calculating sample size for respondent-driven sampling studies of injection drug users in the United States. *AIDS Behav.* **16**, 797–806 (2012).
31. Johnston, L. G., Malekinejad, M., Kendall, C., Iuppa, I. M. & Rutherford, G. W. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. *AIDS Behav.* **12**, S131–141 (2008).
32. Kerr, L. *et al*. HIV prevalence among men who have sex with men in Brazil: results of the 2nd national survey using respondent-driven sampling. *Med.* **97**, S9–S15 (2018).
33. Guimarães, M. D. C. *et al*. Comparing HIV risk-related behaviors between 2 RDS national samples of MSM in Brazil, 2009 and 2016. *Med.* **97**, S62–S68 (2018).

Author contributions

O.R., M.F.C.G. and P.Y.B. designed the simulation part and applied NMA to the Brazilian dataset; C.K. and L.K. were involved in the Brazilian R.D.S. survey; O.R., P.Y.B., M.F.C.G., A.P., C.K. and L.K. analyzed data; O.R., P.Y.B. and C.K. wrote the paper assisted by all other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59567-2.

Correspondence and requests for materials should be addressed to O.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

3.5 Principaux résultats et réflexion sur la qualité des données

Dans l'article présenté plus haut, nous démontrons que la méthode modifiée de Gile, NAestim, permet des estimations de prévalence proches des valeurs réelles et de celles trouvées par la méthode de Gile NA lorsque l'homophilie est inconnue. Ce travail permet ainsi de généraliser le modèle NA aux données les plus fréquemment disponibles dans l'étude de l'épidémie liée au VIH à savoir des études dans lesquelles le statut sérologique des contacts est inconnu.

Sur les données réelles, ce travail confirme les valeurs élevées de la prévalence du VIH chez les HSH brésiliens vivant en zone urbaine. Les nouvelles données RDS confirment ces prévalences (177). Les données des deux enquêtes RDS sur les HSH au Brésil de 2009 et 2019 ont été comparées de manière brut ou en utilisant les méthodes d'estimation historiques (154). Ces deux études ont été difficiles à mener dans certaines villes et les biais liées aux graines, au recrutement différentiel et à la taille du réseau sont probables. Comparer à nouveau ces données à l'aide de l'estimateur développé est d'intérêt dans le cadre d'un suivi longitudinal.

La confrontation des modèles aux données réelles nous a fait nous interroger sur les résultats différents selon les méthodes de calcul utilisées. A notre connaissance, il s'agit du premier travail tentant de comparer des estimations obtenues à partir de différents outils d'estimation et utilisant des données réelles. L'explication des différences observées n'est pas toujours évidente. Les patterns de variations des estimations selon les méthodes nous ont orientés vers des biais qui se sont confirmés à l'analyse des données dans leur contexte. Par exemple, pour la ville d'Itajai, le fait que la méthode RDS-1 trouve une valeur inférieure à la méthode NAestim et que RDS2 trouve une valeur supérieure à celle trouvée par NAestim est compatible avec le problème posé par un échantillon RDS « grand » par rapport à la taille du réseau source. Ceci

est confirmé par les données démographiques, Itajai étant la plus petite ville de l'étude, son échantillon représente plus de 10% de la population cible de cette ville.

Les différences d'estimation ont également nécessité de s'intéresser aux méthodes de diagnostic permettant d'évaluer la qualité d'un jeu de donnée de type RDS (171). Le calcul des longueurs des différentes chaines de chaque jeux de données RDS, l'analyse de la convergence des estimations, peuvent expliquer en partie les différences observées entre les outils d'estimation.

Notre travail propose aussi de nouvelles approches permettant d'évaluer l'importance de l'influence des graines. Un résultat majeur des simulations, qui semble se confirmer sur les données réelles, est que lorsque les graines sont non représentatives, ici par exemple toutes séropositives ou séronégatives, leur influence peut s'étendre sur de nombreuses vagues, plus de 10 dans nos simulations. Cet effet peut être atténué grâce à l'utilisation de la méthode NAEstim que nous proposons. L'outil calculant les estimations en enlevant les vagues une à une en commençant par la première permet de visualiser la convergence plus précocement en minorant l'effet des premières vagues à mesure qu'elles sont enlevées. Elle permet également de maintenir dans l'analyse le plus grand nombre de vague qui ne modifie pas l'estimation.

Enfin, il résulte de ce travail la nécessité d'utiliser plusieurs méthodes d'estimation d'un paramètre à partir d'un même jeu de donnée en vue d'étudier leurs variations mais également d'avoir un regard critique, à partir d'outils diagnostics sur la qualité des données collectées autorisant ou non certaines méthodes d'estimation.

4. Synthèse

Cette thèse est un triptyque permettant d'explorer les différentes utilisations possibles de données de réseaux issues de jeux de données égocentriques et mixtes. L'analyse, la simulation et l'estimation se suivent logiquement, la simulation d'épidémie nécessitant l'analyse en vue de choisir les variables d'intérêt à intégrer dans le modèle épidémique. Et les modèles d'estimation nécessitent des hypothèses sur les variables déterminant les réseaux de travail supposés proche du réseau source.

Les deux premiers volets de cette thèse bénéficient d'une littérature importante et parfois redondante. Les types de variables de réseaux restent les mêmes : homophilie, densité de liens, partenaires simultanés (178). Les variables homophiliques décrites sont souvent les mêmes bien que leur impact puisse être différent selon les populations d'étude (âge, niveau socio-culturel, groupe ethnique). L'analyse des réseaux sociaux utilise donc des variables presque formalisées pour des résultats particuliers. La répétition d'études identiques dans la méthode reste donc absolument nécessaire devant les spécificités des populations clefs au travers le monde. Ces études sont cependant normalisables. Nous voudrions pouvoir intégrer de nouvelles variables de réseaux, plus complexes comme la présence de boucle, de triangle, le calcul de la longueur des chemins entre des individus clefs… L'accès à ces données nécessite de visualiser le réseau de manière quasi-socio centrique. L'effort à consentir est énorme mais nécessaire pour faire avancer la compréhension des phénomènes d'accélération ou d'extinction des phénomènes épidémiques.

La grande question persistante reste la comparaison des jeux de données et donc la comparaison de réseaux issus de populations ou de périodes d'études différentes. Il n'existe pas de travaux comparant différents jeux de données de réseaux. Ceci s'explique par le fait que les données

récoltées utilisant une méthodologie similaire sont rares mais également par le fait que les variables de réseaux sont dépendantes les unes des autres et dépendent de la taille du réseau observé. Un réseau dense aura plus de chance de présenter des structures complexes à type de boucle, de liens simultanés, etc. Ainsi, comparer des variables de réseaux issues de deux réseaux de tailles différentes reste hasardeux.

Les travaux de simulation d'épidémie deviennent de plus en plus complexes grâce aux capacités computationnelles. Ce travail a montré que des réseaux simples peuvent permettre de reproduire des processus épidémiques réels. Le travail méthodologique futur sera de déterminer une méthodologie permettant de valider un modèle de réseau « suffisant » à partir des variables disponibles. Ceci permettra de gagner en comparabilité entre les études et en clarification des messages.

L'évaluation des mesures de prévention en utilisant des modèles individus-centrés deviennent la règle outre-Atlantique et permet de comparer l'efficience de stratégies comme la PrEP au sein de différentes populations à haut risque et aux caractéristiques individuelles et de réseaux différentes (179). En France, la question des mélanges des groupes populationnels d'HSH selon leurs origines géographiques (né en France ou à l'étranger, en Afrique ou ailleurs) ou leur âge se pose devant les différences de prévalence de l'infection par le VIH (180). Ces populations ne sont probablement pas homogènes, tant dans les pratiques sexuelles que dans leur structure du réseau de contact. Ainsi, des travaux sont nécessaires afin d'identifier comment les populations se mélangent et comment elles sont structurées dans leurs contacts sexuels en vue d'évaluer l'efficacité des programmes de préventions par des modèles de simulation individu-centrés.

Les études RDS ont un attrait énorme pour parvenir à décrire les populations ségréguées. De plus, les RDS peuvent avoir un intérêt dans des populations qui sont difficiles d'accès non pas par le fait qu'elles sont rejetées mais par le fait qu'elles ne sont pas atteignables pour un message particulier. Les enquêtes de prévention peuvent ne pas toucher un public cible qui ne se sent pas concerné. Pour exemple, une problématique émergente est l'étude en santé des populations de jeunes adultes qui pourtant présentent des spécificités. Cette population est peu en contact avec le soin. Ainsi, utiliser les RDS chez les jeunes peut revêtir un intérêt dans l'étude de la prévalence des IST mais également dans d'autres domaines d'études comportementales liés à la santé : tabagisme, consommation d'alcool ou de drogues.

5. Ouverture : les réseaux de contacts comme outils clefs de l'action

Paradoxalement et comme dit plus haut, l'étude des réseaux de contacts dans la problématique de l'épidémiologie liée au VIH s'est développée dans le but d'agir sur les variables de réseaux (nombre de partenaires, partenaires simultanés) ou de les utiliser à des fins de recrutement (objectif initial du RDS). Aujourd'hui, ils sont plus observés et analysés qu'utilisés à des fins d'intervention.

Dans de nombreuses situations de mise en place de politique de prévention des ISTs, la méthodologie de sensibilisation et de recrutement repose sur l'information générale et l'information donnée par des individus issus de la communauté. Ces pairs sont sélectionnés de par leur caractère central au sein de leur communauté. Cette sélection peut d'ailleurs être rationalisée par l'analyse des variables de réseaux favorables au recrutement (181). Ces individus sont formés à la prévention et aux protocoles proposés. Ces stratégies sont indispensables mais se limitent aux personnes les plus accessibles de la communauté et aux personnes se sentant les plus concernées. Elles sont par ailleurs couteuses en temps, le recrutement ne dépendant que de peu d'individus.

L'utilisation des réseaux de contacts dans un objectif d'action préventive a été développé ces dix dernières années avec l'avènement des tests rapides à orientation diagnostique (TROD) . Le principe en est que les personnes recrutées dans un programme de prévention ont la possibilité de fournir à leurs partenaires des TROD. Cependant, les personnes ayant bénéficié des TROD ne deviennent pas forcément des recruteurs (182,183). Le réseau atteint n'est ainsi que d'ordre 1 et des chaines de contact ne sont pas mises en évidence. A l'inverse, les études de type RDS à des fins d'épidémiologie descriptive n'ont pas comme objectif la mise en place de stratégies préventives. Cependant, elles proposent aux individus recrutés, au minimum, une orientation vers des organismes proposant un accès à la médecine préventive. Ceci est une

obligation éthique. Curieusement, l'engagement réel d'individus dans une démarche de réduction du risque suite à la participation à ce type d'étude n'a pas été évalué.

Le recrutement en chaine est à nouveau mis au goût du jour depuis peu, surtout sur le plan théorique dans des travaux de modélisation s'intéressant à la mise en place de projet vaccinaux. En cas de découverte d'un vaccin contre le VIH, l'utilisation des RDS serait probablement efficiente en termes de recrutement de populations à haut risque (184,185). L'utilisation de ce type de recrutement dans des essais thérapeutiques est également une piste de travail à développer dans le but d'inclure dans des protocoles des individus jugés difficiles d'accès.

Ainsi, les stratégies de recrutement sont multiples (figure 8), le recrutement par des individus centraux (186) et le recrutement en chaine (187) sont probablement complémentaires. Cette dernière est à développer.

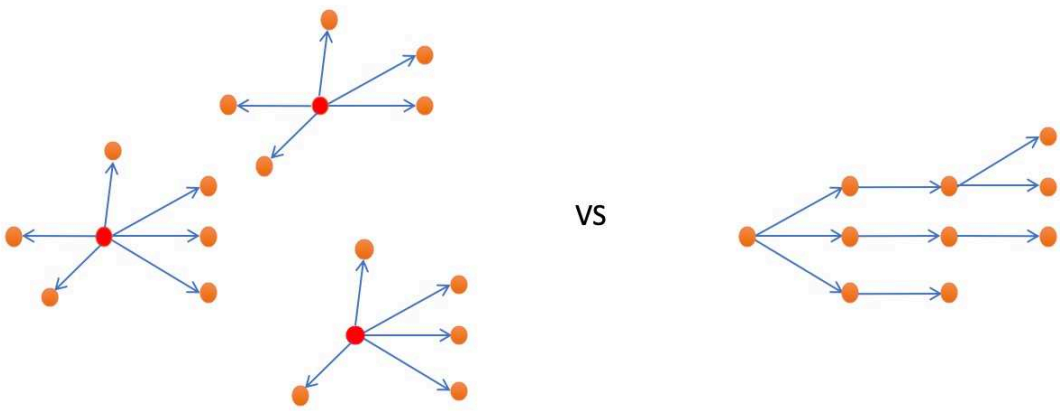

Figure 8 Stratégies de recrutement dans des programmes de prévention.
A gauche, stratégie centrée sur des individus choisis dans la communauté. A droite, stratégie de recrutement en chaine.

Conclusion

Tout au long de ce travail de thèse la question du réseau est restée centrale. L'ensemble des possibilités issu de leur étude nécessite un socle commun de connaissances sur l'historique des méthodes, le vocabulaire et la notion de dépendance entre les quatre éléments qui le structurent que sont les *egos*, les *alters*, leurs liens et les messages.

Ce travail a permis de mieux définir l'impact des réseaux de contacts dans la transmission du VIH au sein des HSH en France et a permis de démontrer qu'à couverture PreP égale, son effet populationnel dépend des variables définissant le réseau de contact sexuel.

La modélisation de réseau à des fins d'estimation de prévalence à partir de donnés RDS a démontré sa supériorité comparativement aux méthodes antérieures. Notre méthode rend cet outil d'estimation utilisable sur des jeux de données réelles.

Ces travaux ont été réalisés à partir de données initialement non récoltées dans ce but ou issues d'un autre pays rompu à la réalisation de ce type d'étude, le Brésil. Cela souligne l'importance en France de développer des projets de recherche sur les réseaux de contacts sexuels.

Enfin, l'action débute par l'observation, ces travaux non interventionnels ont naturellement fait aboutir la réflexion sur l'utilisation des propriétés de réseau dans l'optimisation des politiques de prévention.

Bibliographie

1. Anderson RM, Anderson B, May RM. Infectious Diseases of Humans: Dynamics and Control. OUP Oxford; 1992. 772 p.
2. Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. Soc Sci Med. 1985;21(11):1203–16.
3. Gupta S, Anderson RM, May RM Networks of Sexual Contacts: Implications for the Pattern of Spread of HIV Vol. 3, AIDS (London, England). AIDS; 1989 Dec;3(12):807-17.
4. Vespignani A Twenty Years of Network Science [Internet]. Vol. 558, Nature. 558, 528-529 (2018)
5. Morris M Network Epidemiology Oxford University Press 2004
6. Guzzetta G, Ajelli M, Yang Z, Merler S, Furlanello C, Kirschner D. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. J Theor Biol. 2011 Nov 21;289:197–205.
7. Cohen T, Colijn C, Finklea B, Murray M. Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network model of transmission. J R Soc Interface. 2007 Jun 22;4(14):523–31.
8. Barniol J, Niemann S, Louis VR, Brodhun B, Dreweck C, Richter E, et al. Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant sub-populations. BMC Infect Dis. 2009 Dec 4;9:197.
9. Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of infectious disease. Math Biosci. 1996 Apr 15;133(2):165–95.
10. Bogunia M, Pastor-Satorras R. Epidemic Spreading in Correlated Complex Networks Phys Rev E Stat Nonlin Soft Matter Phys; 2002 Oct;66(4 Pt 2)
11. Gile KJ, Handcock MS. Network Model-Assisted Inference from Respondent-Driven Sampling Data. J R Stat Soc Ser A Stat Soc. 2015 Jun;178(3):619–39.
12. Le Vu S, Le Strat Y, Barin F, Pillonel J, Cazein F, Bousquet V, et al. Population-based HIV-1 incidence in France, 2003-08: a modelling analysis. Lancet Infect Dis. 2010 Oct;10(10):682–7.
13. Fiche d'information 2019 — Dernières statistiques sur l'état de l'épidémie de sida Available from: https://www.unaids.org/fr/resources/fact-sheet
14. SPF. Bulletin de santé publique VIH/sida. Octobre 2019. [Internet]
15. Sattenspiel L, Koopman J, Simon C, Jacquez JA The Effects of Population Structure on the Spread of the HIV Infection Am J Phys Anthropol; 1990 Aug;82(4):421-9.
16. Adimora AA, Schoenbach VJ. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. J Infect Dis. 2005 Feb 1;191 Suppl 1:S115-122.
17. Helleringer S, Kohler H-P, Chimbiri A, Chatonda P, Mkandawire J. The Likoma Network Study: Context, data collection, and initial results. Demogr Res. 2009;21:427–68.
18. Jenness SM, Goodreau SM, Rosenberg E, Beylerian EN, Hoover KW, Smith DK, et al. Impact of the Centers for Disease Control's HIV Preexposure Prophylaxis Guidelines for Men Who Have Sex With Men in the United States. J Infect Dis. 2016 Jul 14;
19. Rapoport A. Spread of information through a population with socio-structural bias: I. Assumption of transitivity. Bulletin of Mathematical Biophysics. 1953 Dec 1;15(4):523–33.
20. Watts DJ. Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton University Press; 2018.
21. Briton T, Trapman P. Inferring Global Network Properties From Egocentric Data With Applications to Epidemics Math Med Biol; 2015 Mar;32(1):99-112.
22. Heckathorn DD. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. Soc Probl. 1997 May 1;44(2):174–99.

23. Christakis NA, Fowler JH. Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior. Statistics in medicine. 2013 Feb 20;32(4):556-77
24. Hill AL , Rand DG , Nowak MA, Christakis NA Infectious Disease Modeling of Social Contagion in Networks. Vol. 6, PLoS computational biology. PLoS Comput Biol; 2010 Nov 4;6(11)
25. Perkins JM, Subramanian SV, Christakis NA. Social networks and health: a systematic review of sociocentric network studies in low- and middle-income countries. Soc Sci Med. 2015 Jan;125:60–78.
26. Miller JC, Kiss IZ. Epidemic spread in networks: Existing methods and current challenges. Math Model Nat Phenom. 2014 Jan;9(2):4–42.
27. Rothenberg RB, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, Klovdahl AS. Social network dynamics and HIV transmission. AIDS. 1998 Aug 20;12(12):1529–36.
28. Hallfors DD, Iritani BJ, Miller WC, Bauer DJ. Sexual and drug behavior patterns and HIV and STD racial disparities: the need for new directions. Am J Public Health. 2007 Jan;97(1):125–32.
29. Hamilton DT, Morris M. The racial disparities in STI in the U.S.: Concurrency, STI prevalence, and heterogeneity in partner selection. Epidemics. 2015 Jun;11:56–61.
30. Boily M-C, Alary M, Baggaley RF. Neglected Issues and Hypotheses Regarding the Impact of Sexual Concurrency on HIV and Sexually Transmitted Infections. AIDS Behav. 2011 Jan 30
31. Doherty IA, Shiboski S, Ellen JM, Adimora AA, Padian NS. Sexual bridging socially and over time: a simulation model exploring the relative effects of mixing and concurrency on viral sexually transmitted infection transmission. Sex Transm Dis. 2006 Jun;33(6):368–73.
32. adams jimi, Moody J, Morris M. Sex, Drugs, and Race: How Behaviors Differentially Contribute to the Sexually Transmitted Infection Risk Network Structure. Am J Public Health. 2013 Feb;103(2):322–9.
33. Gorbach PM, Weiss RE, Jeffries R, Javanbakht M, Drumright LN, Daar ES, et al. Behaviors of recently HIV-infected men who have sex with men in the year postdiagnosis: effects of drug use and partner types. J Acquir Immune Defic Syndr. 2011 Feb;56(2):176–82.
34. Williams JK, Wilton L, Magnus M, Wang L et al Sexual Networks and HIV Risk Among Black Men Who Have Sex With Men in 6 U.S. Cities [Internet]. Vol. 10, PloS one. PLoS One; 2015 Dec;105(12):2473-81
35. Schechter MT, Craib KJ, Willoughby B, Douglas B, McLeod WA, Maynard M, et al. Patterns of sexual behavior and condom use in a cohort of homosexual men. Am J Public Health. 1988 Dec;78(12):1535–8.
36. Evans BA, McCormack SM, Bond RA, MacRae KD, Thorp RW. Human immunodeficiency virus infection, hepatitis B virus infection, and sexual behaviour of women attending a genitourinary medicine clinic. Br Med J (Clin Res Ed). 1988 Feb 13;296(6620):473–5.
37. Pickering H, Okongo M, Nnalusiba B, Bwanika K, Whitworth J. Sexual networks in Uganda: casual and commercial sex in a trading town. AIDS Care. 1997 Apr;9(2):199–207.
38. Weinberg MS, Lottes IL, Aveline D. AIDS risk reduction strategies among United States and Swedish heterosexual university students. Arch Sex Behav. 1998 Aug;27(4):385–401.
39. Eaton LA, West TV, Kenny DA, Kalichman SC. HIV Transmission Risk among HIV Seroconcordant and Serodiscordant Couples: Dyadic Processes of Partner Selection. AIDS Behav. 2009 Apr;13(2):185–95.
40. Schensul JJ, Levy JA, Disch WB. Individual, contextual, and social network factors affecting exposure to HIV/AIDS risk among older residents living in low-income senior housing complexes. J Acquir Immune Defic Syndr. 2003 Jun 1;33 Suppl 2:S138-152.

41. Bouhnik A-D, Préau M, Schiltz M-A, Lert F, Obadia Y, Spire B, et al. Unprotected sex in regular partnerships among homosexual men living with HIV: a comparison between sero-nonconcordant and seroconcordant couples (ANRS-EN12-VESPA Study). AIDS. 2007 Jan;21 Suppl 1:S43-48.
42. Lavoie E, Alary M, Remis RS, Otis J, Vincelette J, Turmel B, et al. Determinants of HIV seroconversion among men who have sex with men living in a low HIV incidence population in the era of highly active antiretroviral therapies. Sex Transm Dis. 2008 Jan;35(1):25–9.
43. de Walque D, Kline R. Variations in condom use by type of partner in 13 sub-Saharan African countries. Stud Fam Plann. 2011 Mar;42(1):1–10.
44. Goodreau S, Cassels S, Kasprzyk D, Montaño D, Greek A, Morris M. Concurrent Partnerships, Acute Infection and HIV Epidemic Dynamics Among Young Adults in Zimbabwe. AIDS and Behavior. 2012;16(2):312–22.
45. Prestage G, Jin F, Bavinton B, Scott SA, Hurley M. Do differences in age between sexual partners affect sexual risk behaviour among Australian gay and bisexual men? Sex Transm Infect. 2013 Dec;89(8):653–8.
46. Newcomb ME, Mustanski B Racial Differences in Same-Race Partnering and the Effects of Sexual Partnership Characteristics on HIV Risk in MSM: A Prospective Sexual Diary Study J Acquir Immune Defic Syndr; 2013 Mar 1;62(3):329-33.
47. Mustanski B, Newcomb ME, Clerkin E. Relationship Characteristics and Sexual Risk-Taking in Young Men Who Have Sex With Men. Health Psychol; 2011 Sep; 30(5): 597–605.
48. Sullivan PS, Salazar L, Buchbinder S, Sanchez TH. Estimating the Proportion of HIV Transmissions From Main Sex Partners Among Men Who Have Sex With Men in Five US Cities [Internet]. Vol. 23, AIDS (London, England). AIDS; 2009 Jun 1;23(9):1153-62
49. Grov C, Parsons JT, Bimbi DS. Sexual risk behavior and venues for meeting sex partners: an intercept survey of gay and bisexual men in LA and NYC. AIDS Behav. 2007 Nov;11(6):915–26.
50. Grov C, Rendina HJ, Ventuneac A, Parsons JT. Partners met via sex parties present significantly greater odds for condomless anal sex among MSM: an event-level analysis of venues where male partners are met. J Acquir Immune Defic Syndr. 2014 Dec 15;67(5):564–8.
51. Grov C, Crow T. Attitudes about and HIV risk related to the "most common place" MSM meet their sex partners: comparing men from bathhouses, bars/clubs, and Craigslist.org. AIDS Educ Prev. 2012 Apr;24(2):102–16.
52. Brantley M, Schumacher C, Fields EL, Perin J, Safi AG, Ellen JM, et al. The network structure of sex partner meeting places reported by HIV-infected MSM: Opportunities for HIV targeted control. Soc Sci Med. 2017;182:20–9.
53. Jennings JM, Tilchin C , Meza B, Schumacher C et al. Overlapping Transmission Networks of Early Syphilis and/or Newly HIV Diagnosed Gay, Bisexual and Other Men Who Have Sex With Men (MSM): Opportunities for Optimizing Public Health Interventions [Internet]. AIDS and behavior. AIDS Behav; 2020 Apr 1
54. HIV: consensus indicators are needed for concurrency. Lancet. 2010 Feb 20;375(9715):621–2.
55. Aral SO. Partner Concurrency and the STD/HIV Epidemic. Curr Infect Dis Rep. 2010 Mar;12(2):134–9.
56. Garnett GP, Johnson AM. Coining a new term in epidemiology: concurrency and HIV. AIDS. 1997 Apr;11(5):681–3.
57. Goodreau SM. Is 2 a "high number of partners"? Modeling, data, and the power of concurrency. Sex Transm Dis. 2013 Jan;40(1):61.
58. Maulsby C, Jain K, Sifakis F, German D, Flynn CP, Holtgrave D. Individual-Level and

Partner-Level Predictors of Newly Diagnosed HIV Infection Among Black and White Men Who Have Sex with Men in Baltimore, MD. AIDS Behav. 2015 May 1;19(5):909–17.
59. Tieu H-V, Nandi V, Hoover DR, Lucy D, Stewart K, Frye V, et al. Do Sexual Networks of Men Who Have Sex with Men in New York City Differ by Race/Ethnicity? AIDS Patient Care STDS. 2016 Jan 1;30(1):39–47.
60. Brown RE, Turner C, Hern J, Santos GM Partner-level Substance Use Associated With Increased Sexual Risk Behaviors Among Men Who Have Sex With Men in San Francisco, CA . Drug Alcohol Depend; 2017 Jul 1;176:176-180.
61. Grieb SMD, Davey-Rothwell M, Latkin CA. Social and Sexual Network Characteristics and Concurrent Sexual Partnerships Among Urban African American High-risk Women with Main Sex Partners. AIDS and Behavior. 2011 Aug 23
62. Tobin K, Davey-Rothwell M, Latkin C. Social-Level Correlates of Shooting Gallery Attendance: a focus on networks and norms. AIDS Behav. 2010 Oct;14(5):1142–8.
63. Latkin CA, Tieu HV, Fields S, Hanscom BS, Connor M, Hanscom B, et al. Social Network Factors as Correlates and Predictors of High Depressive Symptoms Among Black Men Who Have Sex with Men in HPTN 061. AIDS Behav. 2017 Apr 1;21(4):1163–70.
64. Friedman SR, Neaigus A, Joe B, Curtis R, Goldstein M, Ildefonso G, et al. Sociometric Risk Networks and Risk for HIV Infection. Vol. 87, American journal of public health. Am J Public Health; 1997 Aug 1289-1296
65. Neaigus A, Friedman SR, Goldstein M, Ildefonso G, Curtis R, Jose B. Using dyadic data for a network analysis of HIV infection and risk behaviors among injecting drug users. NIDA Res Monogr. 1995;151:20–37.
66. Gyarmathy VA, Neaigus A. The Effect of Personal Network Exposure on Injecting Equipment Sharing among IDUs in Budapest, Hungary. : CONNECTIONS 2004 15(1): 29-42
67. Weeks MR, Clair S, Borgatti SP, Radda K, Schensul JJ. Social Networks of Drug Users in High-Risk Sites: Finding the Connections. AIDS Behav. 2002 Jun 1;6(2):193–206.
68. Kottiri BJ, Friedman SR, Neaigus A, Curtis R, Des Jarles DC. Risk Networks and Racial/Ethnic Differences in the Prevalence of HIV Infection Among Injection Drug Users J Acquir Immune Defic Syndr; 2002 May 1;30(1):95-104
69. Shehan DM, Miller RP, Trepka MJ, Smith LR, Latkin C. Role of Social Network Sexual Norms and Behaviors on the HIV Sexual Risk Behaviors of People Who Inject Drugs in HPTN 037 AIDS and behavior. AIDS Behav; 2019 Jun;23(6):1604-1611
70. Pines HA, Wertheim JO, Liu L, Garfein RS, Little SJ, Karris MY. Concurrency and HIV transmission network characteristics among MSM with recent HIV infection. AIDS. 2016 Nov 28;30(18):2875–83.
71. Hicks MR, Kogan SM, Cho J, Oshri A. Condom Use in the Context of Main and Casual Partner Concurrency: Individual and Relationship Predictors in a Sample of Heterosexual African American Men. Am J Mens Health. 2016 May 18;
72. Bello B, Moultrie H , Somji A , Chersich MF , Watts C, Delany-Moretlwe S Alcohol Use and Sexual Risk Behaviour Among Men and Women in Inner-City Johannesburg, South Africa BMC Public Health; 2017 Jul 4;17(Suppl 3):548
73. Kenyon CR, Tsoumanis A , Schwartz IS , Maughan-Brown B Partner Concurrency and HIV Infection Risk in South Africa Int J Infect Dis; 2016 Apr;45:81-7.
74. Périssé ARS, Langenberg P, Hungerford L, Boulay M, Charurat M, Schechter M, et al. Egocentric network data provide additional information for characterizing an individual's HIV risk profile. AIDS. 2010 Jan 16;24(2):291–8.
75. Anderson RM: Infectious diseases of humans: Dynamics and Control 1992 Oxford Science Publication
76. Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res. 1993 Mar 1;2(1):23–41.

77. Macdonald G. The Analysis of Equilibrium in Malaria Vol. 49, Tropical diseases bulletin. Trop Dis Bull; 1952
78. May RM, Anderson RM. Transmission dynamics of HIV infection. Nature. 1987 Mar 12;326(6109):137–42.
79. May RM, Anderson RM. The Transmission Dynamics of Human Immunodeficiency Virus (HIV) Philos Trans R Soc Lond B Biol Sci; 1988
80. Anglemyer A, Rutherford GW, Egger M, Siegfried N. Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples. Cochrane Database Syst Rev. 2011;5:CD009153.
81. Diekmann O, Heesterbeek J a. P, Metz J a. J. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. J Math Biol. 1990 Jun 1;28(4):365–82.
82. Anderson RM. The Epidemiology of HIV Infection: Variable Incubation Plus Infectious Periods and Heterogeneity in Sexual Activity. Journal of the Royal Statistical Society Series A (Statistics in Society). 1988;151(1):66–98.
83. Jacquez JA, Simon CP, Koopman J, Sattenspiel L, Perry T. Modeling and analyzing HIV transmission: the effect of contact patterns. Mathematical Biosciences. 1988 Dec 1;92(2):119–99.
84. Anderson RM. Some Aspects of Sexual Behaviour and the Potential Demographic Impact of AIDS in Developing Countries Soc Sci Med; 1992 Feb;34(3):271-80.
85. Blythe SP, Castillo-Chavez C. Like-with-like Preference and Sexual Mixing Models Math Biosci; 1989 96 (2) 221-238
86. Dietz K. On the transmission dynamics of HIV. Mathematical Biosciences. 1988 Jul;90(1–2):397–414.
87. Dietz K, Hadeler KP. Epidemiological models for sexually transmitted diseases. Journal of Mathematical Biology. 1988;26(1):1–25.
88. Dietz K, D Schenzle D. Proportionate Mixing Models for Age-Dependent Infection Transmission. J Math Biol; 1985;22(1):117-20.
89. Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. Impact of Sampling Density on the Extent of HIV Clustering. AIDS Res Hum Retroviruses. 2014 Oct 2;
90. Hamilton DT, Handcock MS, Morris M. Degree distributions in sexual networks: a framework for evaluating evidence. Sex Transm Dis. 2008 Jan;35(1):30–40.
91. Janulis P, Phillips G, Birkett M, Mustanski B. Sexual Networks of Racially Diverse Young MSM Differ in Racial Homophily But Not Concurrency. J Acquir Immune Defic Syndr. 2018 15;77(5):459–66.
92. Kenyon C, Colebunders R. Birds of a feather: homophily and sexual network structure in sub-Saharan Africa. Int J STD AIDS. 2013 Mar 27;
93. Shalizi CR, Thomas AC. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociol Methods Res. 2011 May;40(2):211–39.
94. Morris M. Put in Context: The Concurrency Intervention Simulation Study of McCreesh et al. Sex Transm Dis. 2012 Oct;39(12):993–4.
95. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. Am J Public Health. 2009 Jun;99(6):1023–31.
96. Carnegie NB, Morris M. Size Matters: Concurrency and the Epidemic Potential of HIV in Small Networks. PLoS One. 2012 Aug 24
97. Morris M, Epstein H, Wawer M. Timing is everything: international variations in historical sexual partnership concurrency and HIV prevalence. PLoS ONE. 2010;5(11):e14092.

98. Goodreau SM, Rosenberg ES, Jenness SM, Luisi N, Stansfield SE, Millett GA, et al. Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study. Lancet HIV. 2017 Jul;4(7):e311-e320.
99. Berge C. Two Theorems in Graph Theory. PNAS. 1957 Sep 15;43(9):842–4.
100. Newman M. Networks [Internet]. Oxford University Press; 2010
101. On Random Graphs I | BibSonomy [Internet]. [cited 2020 May 25]. Available from: https://www.bibsonomy.org/bibtex/99061fc859ba540d4485abfbce44f298
102. Barabasi AL, Albert R. Emergence of Scaling in Random Networks [Internet]. Vol. 286, Science (New York, N.Y.). Science; 15 Oct 1999:Vol. 286, Issue 5439, pp. 509-512
103. Frank O, Strauss D. Markov Graphs. Journal of the American Statistical Association. 1986;81(395):832–42.
104. Goodreau SM. Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network. Soc Networks. 2007 May;29(2):231–48.
105. Krivitsky PN. Modeling Tie Duration in ERGM-Based Dynamic Network Models.
106. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. J Stat Softw. 2008 May 1;24(3):nihpa54860.
107. Morris M, Handcock MS, Hunter DR. Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. J Stat Softw. 2008;24(4):1548–7660.
108. Desmarais BA, Cranmer SJ. Statistical inference for valued-edge networks: the generalized exponential random graph model. PLoS ONE. 2012;7(1):e30136.
109. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. J Stat Softw. 2008;24(1):1548–7660.
110. Krivitsky PN. Package examples for networkDynamic: Dynamic Extensions for Network Objects (Version 0.2-2).
111. Krivitsky P. STERGM - Separable Temporal ERGMs for modeling discrete relational dynamics with statnet.
112. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis. 2005 May 1;191(9):1403–9.
113. Pinkerton SD. Probability of HIV transmission during acute infection in Rakai, Uganda. AIDS Behav. 2008 Sep;12(5):677–84.
114. Abu-Raddad LJ, Longini IM. No HIV stage is dominant in driving the HIV epidemic in sub-Saharan Africa. AIDS. 2008 May 31;22(9):1055–61.
115. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. J Infect Dis. 2008 Sep 1;198(5):687–93.
116. Goodreau SM, Cassels S, Kasprzyk D, Montaño DE, Greek A, Morris M. Concurrent Partnerships, Acute Infection and HIV Epidemic Dynamics Among Young Adults in Zimbabwe. AIDS Behav. 2010 Dec 30
117. Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of infectious disease. Math Biosci. 1996 Apr 15;133(2):165–95.
118. Jenness SM, Goodreau SM, Morris M, Cassels S. Effectiveness of combination packages for HIV-1 prevention in sub-Saharan Africa depends on partnership network structure: a mathematical modelling study. Sex Transm Infect. 2016 Jun 9;
119. Annie V, Francis B, Stephan LV, Josiane P, et al. Prévalence du VIH et comportement de dépistage des hommes fréquentant les lieux de convivialité gay parisiens, Prevagay 2009. Bull Epidemiol Hebd. 2010;464–7.
120. SPF. Rapport Prevagay 2015. Enquête de séroprévalence du VIH menée auprès des hommes ayant des relations sexuelles avec des hommes fréquentant les lieux de convivialité

gay
121. Røttingen J-A, Garnett GP. The epidemiological and control implications of HIV transmission probabilities within partnerships. Sex Transm Dis. 2002 Dec;29(12):818–27.
122. Wilson DP, Law MG, Grulich AE, Cooper DA, Kaldor JM. Relation between HIV viral load and infectiousness: a model-based analysis. Lancet. 2008 Jul 26;372(9635):314–20.
123. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. Cochrane Database Syst Rev. 2001;(3):CD003255.
124. Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, et al. Antiretroviral Preexposure Prophylaxis for Heterosexual HIV Transmission in Botswana. The New England journal of medicine [Internet]. 2012 Jul 11
125. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, et al. Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. The New England journal of medicine [Internet]. 2012 Jul 11
126. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. New England Journal of Medicine. 2010 Dec 30;363(27):2587–99.
127. Phillips A, Pezzotti P. Short-term risk of AIDS according to current CD4 cell count and viral load in antiretroviral drug-naive individuals and those treated in the monotherapy era. AIDS. 2004 Jan 2;18(1):51–8.
128. Phillips AN, Lundgren JD. The CD4 lymphocyte count and risk of clinical progression. Curr Opin HIV AIDS. 2006 Jan;1(1):43–9.
129. Lyles RH, Muñoz A, Yamashita TE, Bazmi H, Detels R, Rinaldo CR, et al. Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. J Infect Dis. 2000 Mar;181(3):872–80.
130. Lyles CM, Dorrucci M, Vlahov D, Pezzotti P, Angarano G, Sinicco A, et al. Longitudinal Human Immunodeficiency Virus Type 1 Load in the Italian Seroconversion Study: Correlates and Temporal Trends of Virus Load. J Infect Dis. 1999 Jan 10;180(4):1018–24.
131. Hubert JB, Burgard M, Dussaix E, Tamalet C, Deveau C, Le Chenadec J, et al. Natural history of serum HIV-1 RNA levels in 330 patients with a known date of infection. The SEROCO Study Group. AIDS. 2000 Jan 28;14(2):123–31.
132. Tarwater PM, Gallant JE, Mellors JW, Gore ME, Phair JP, Detels R, et al. Prognostic value of plasma HIV RNA among highly active antiretroviral therapy users. AIDS. 2004 Dec 3;18(18):2419–23.
133. García-Lerma JG, Otten RA, Qari SH, Jackson E, Cong M, Masciotra S, et al. Prevention of Rectal SHIV Transmission in Macaques by Daily or Intermittent Prophylaxis with Emtricitabine and Tenofovir. PLoS Med [Internet]. 2008 Feb
134. Supervie V, García-Lerma JG, Heneine W, Blower S. HIV, transmitted drug resistance, and the paradox of preexposure prophylaxis. Proc Natl Acad Sci U S A. 2010 Jul 6;107(27):12381–6.
135. Madge S, Smith CJ, Lampe F, Sabin CA, Youle M, Johnson MA, et al. An audit of viral load in one clinical population to describe features of viraemic patients on antiretroviral therapy. HIV Med. 2008 Apr;9(4):208–13.
136. Pilcher CD, Tien HC, Eron JJ, Vernazza PL, Leu S-Y, Stewart PW, et al. Brief but Efficient: Acute HIV Infection and the Sexual Transmission of HIV. J Infect Dis. 2004 May 15;189(10):1785–92.
137. Pilcher CD, Joaki G, Hoffman IF, Martinson FEA, Mapanje C, Stewart PW, et al. Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection. AIDS. 2007 Aug 20;21(13):1723–30.

138. Kaufmann GR, Cunningham P, Kelleher AD, Zaunders J, Carr A, Vizzard J, et al. Patterns of Viral Dynamics During Primary Human Immunodeficiency Virus Type 1 Infection. J Infect Dis. 1998 Jan 12;178(6):1812–5.
139. Meunier E, Siegel K. Sex Club/Party Attendance and STI Among Men Who Have Sex With Men: Results From an Online Survey in New York City. Sex Transm Infect; 2019 Dec;95(8):584-587.
140. Grov C, Hirshfield S, Remien RH, Humberstone M, Chiasson MA Exploring the Venue's Role in Risky Sexual Behavior Among Gay and Bisexual Men: An Event-Level Analysis From a National Online Survey in the U.S Arch Sex Behav; 2013 Feb; 42(2): 291–302.
141. Lankowski A, Sanchez H,Nidalgo J, Cabello R, Duer A Sex-on-premise Venues, Associated Risk Behaviors, and Attitudes Toward Venue-Based HIV Testing Among Men Who Have Sex With Men in Lima, Perú [Internet]. BMC Public Health; 2020; 20: 521
142. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med. 2010 Dec 30;363(27):2587–99.
143. Molina J-M, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, et al. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. N Engl J Med. 2015 Dec 3;373(23):2237–46.
144. Koester KA , Erguera XA , Dufour MSK, Udoh I, Burack JH , Grant RM, Myers J "Losing the Phobia:" Understanding How HIV Pre-exposure Prophylaxis Facilitates Bridging the Serodivide Among Men Who Have Sex With Men. Front Public Health; 2018 Sep 6;6:250
145. Broadhead RS, Heckathorn DD. AIDS Prevention Outreach Among Injection Drug Users: Agency Problems and New Approaches. Soc Probl. 1994 Aug 1;41(3):473–95.
146. Broadhead RS, Heckathorn DD, Weakliem DL, Anthony DL, Madray H, Mills RJ, et al. Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention. Public Health Rep. 1998 Jun;113 Suppl 1:42–57.
147. Heckathorn DD, Broadhead RS, Anthony DL, Weakliem DL. Aids and Social Networks: HIV Prevention Through Network Mobilization. Sociological Focus. 1999 May;32(2):159–79.
148. Goodman LA. Snowball Sampling. Ann Math Statist. 1961 Mar;32(1):148–70.
149. WHO, Introduction to HIV/AIDS and sexually transmitted infection surveillance: module 4: introduction to respondent-driven sampling. 2013
150. White RG, Hakim AJ, Salganik MJ, Spiller MW, Johnston LG, Kerr L, et al. Strengthening the Reporting of Observational Studies in Epidemiology for respondent-driven sampling studies: "STROBE-RDS" statement. J Clin Epidemiol. 2015 May 1;
151. Bazazi AR, Crawford F, Zelenev A, Heimer R, Kamarulzaman A, Altice FL. HIV Prevalence Among People Who Inject Drugs in Greater Kuala Lumpur Recruited Using Respondent-Driven Sampling. AIDS Behav. 2015 Dec;19(12):2347–57.
152. Wejnert C, Pham H, Krishna N, Le B, DiNenno E. Estimating design effect and calculating sample size for respondent-driven sampling studies of injection drug users in the United States. AIDS Behav. 2012 May;16(4):797–806.
153. Heckathorn DD, Semaan S, Broadhead RS, Hughes JJ. Extensions of Respondent-Driven Sampling: A New Approach to the Study of Injection Drug Users Aged 18–25. AIDS Behav. 2002 Mar 1;6(1):55–67.
154. Guimarães MDC, Kendall C, Magno L, Rocha GM, Knauth DR, Leal AF, et al. Comparing HIV risk-related behaviors between 2 RDS national samples of MSM in Brazil, 2009 and 2016. Medicine (Baltimore). 2018 May;97(1S Suppl 1):S62–8.
155. Carballo-Diéguez A, Balan I, Marone R, Pando MA, Dolezal C, Barreda V, et al. Use of respondent driven sampling (RDS) generates a very diverse sample of men who have sex

with men (MSM) in Buenos Aires, Argentina. PLoS ONE. 2011;6(11):e27447.
156. Lane T, Raymond HF, Dladla S, Rasethe J, Struthers H, McFarland W, et al. High HIV prevalence among men who have sex with men in Soweto, South Africa: results from the Soweto Men's Study. AIDS Behav. 2011 Apr;15(3):626–34.
157. Lauby JL, Millett GA, LaPollo AB, Bond L, Murrill CS, Marks G. Sexual risk behaviors of HIV-positive, HIV-negative, and serostatus-unknown Black men who have sex with men and women. Arch Sex Behav. 2008 Oct;37(5):708–19.
158. Johnston LG, Malekinejad M, Kendall C, Iuppa IM, Rutherford GW. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. AIDS Behav. 2008 Jul;12(4 Suppl):S131-141.
159. Kerr LRFS, Mota RS, Kendall C, Pinho A de A, Mello MB, Guimarães MDC, et al. HIV among MSM in a large middle-income country. AIDS. 2013 Jan 28;27(3):427–35.
160. Jalil EM, Wilson EC, Luz PM, Velasque L, Moreira RI, Castro CV, et al. HIV testing and the care continuum among transgender women: population estimates from Rio de Janeiro, Brazil. J Int AIDS Soc. 2017 19;20(1):21873.
161. Bauer GR, Travers R, Scanlon K, Coleman TA. High heterogeneity of HIV-related sexual risk among transgender people in Ontario, Canada: a province-wide respondent-driven sampling survey. BMC Public Health. 2012 Apr 20;12:292.
162. Kriitmaa K, Testa A, Osman M, Bozicevic I, Riedner G, Malungu J, et al. HIV prevalence and characteristics of sex work among female sex workers in Hargeisa, Somaliland, Somalia. AIDS. 2010 Jul;24 Suppl 2:S61-67.
163. Liu H, Liu H, Cai Y, Rhodes AG, Hong F. Money boys, HIV risks, and the associations between norms and safer sex: a respondent-driven sampling study in Shenzhen, China. AIDS Behav. 2009 Aug;13(4):652–62.
164. Johnston LG, Sabin K, Mai TH, Pham TH. Assessment of respondent driven sampling for recruiting female sex workers in two Vietnamese cities: reaching the unseen sex worker. J Urban Health. 2006 Nov;83(6 Suppl):i16-28.
165. Gyarmathy VA, Johnston LG, Caplinskiene I, Caplinskas S, Latkin CA. A simulative comparison of respondent driven sampling with incentivized snowball sampling – the "strudel effect." Drug Alcohol Depend. 2014 Feb 1;135:71–7.
166. Wei C, McFarland W, Colfax GN, Fuqua V, Raymond HF. Reaching black men who have sex with men: a comparison between respondent-driven sampling and time-location sampling. Sex Transm Infect. 2012 Dec;88(8):622–6.
167. Platt L, Wall M, Rhodes T, Judd A, Hickman M, Johnston LG, et al. Methods to recruit hard-to-reach groups: comparing two chain referral sampling methods of recruiting injecting drug users across nine studies in Russia and Estonia. J Urban Health. 2006 Nov;83(6 Suppl):i39-53.
168. Kral AH, Malekinejad M, Vaudrey J, Martinez AN, Lorvick J, McFarland W, et al. Comparing respondent-driven sampling and targeted sampling methods of recruiting injection drug users in San Francisco. J Urban Health. 2010 Sep;87(5):839–50.
169. Robinson WT, Risser JMH, McGoy S, Becker AB, Rehman H, Jefferson M, et al. Recruiting injection drug users: a three-site comparison of results and experiences with respondent-driven and targeted sampling procedures. J Urban Health. 2006 Nov;83(6 Suppl):i29-38.
170. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS. 2005 May;19 Suppl 2:S67-72.
171. Gile KJ, Johnston LG, Salganik MJ. Diagnostics for Respondent-driven Sampling. Journal of the Royal Statistical Society Series A, (Statistics in Society). 2015 Jan;178(1):241.
172. Salganik MJ, Heckathorn DD. Sampling and Estimation in Hidden Populations Using

Respondent-Driven Sampling. Sociological Methodology. 2004;34:193–239.
173. Heckathorn DD. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. Social Problems. 2002;49(1):11–34.
174. Volz E, Heckathorn DD. Probability Based Estimation Theory for Respondent Driven Sampling. J Off Stat. 2008;24(1):79–97.
175. Gile KJ, Handcock MS. Respondent-Driven Sampling: An Assessment of Current Methodology. Sociol Methodol. 2010 Aug;40(1):285–327.
176. Gile KJ. Improved Inference for Respondent-Driven Sampling Data with Application to HIV Prevalence Estimation. arXiv:10064837 [stat] [Internet]. 2010 Jun 24
177. Kerr L, Kendall C, Guimarães MDC, Salani Mota R, Veras MA, Dourado I, et al. HIV prevalence among men who have sex with men in Brazil: results of the 2nd national survey using respondent-driven sampling. Medicine (Baltimore). 2018 May;97(1S Suppl 1):S9–15.
178. Schmid BV, Kretzschmar M. Determinants of Sexual Network Structure and Their Impact on Cumulative Network Measures. PLoS Comput Biol. 2012 Apr;8(4).
179. Goedel WC, Bessey S, Lurie MN, Biello KB, Sullivan PS, Nunn AS, et al. Projecting the impact of equity-based pre-exposure prophylaxis implementation on racial disparities in HIV incidence among men who have sex with men. AIDS. 2020 May 6;
180. Marty L, Cazein F, Panjo H, Pillonel J, Costagliola D, Supervie V, et al. Revealing geographical and population heterogeneity in HIV incidence, undiagnosed HIV prevalence and time to diagnosis to improve prevention and care: estimates for France. J Int AIDS Soc. 2018;21(3):e25100.
181. Walsh T, Schneider JA, Ardestani BM, Young LE. Individual and Social Network Structure Characteristics Associated with Peer Change Agent Engagement and Impact in a PrEP Intervention. AIDS Behav. 2020 May 11;
182. Lippman SA, Lane T, Rabede O, Gilmore H, Chen Y-H, Mlotshwa N, et al. High Acceptability and Increased HIV-Testing Frequency After Introduction of HIV Self-Testing and Network Distribution Among South African MSM. J Acquir Immune Defic Syndr. 2018 01;77(3):279–87.
183. Xiao W, Yan L, Chen L, Fu G, Yang H, Yang C, et al. Sexual network distribution of HIV self-testing kits: Findings from the process evaluation of an intervention for men who have sex with men in China. PLoS ONE. 2020;15(4):e0232094.
184. Chen S, Lu X. An Immunization Strategy for Hidden Populations. Sci Rep. 2017 12;7(1):3268.
185. Mbunda T, Tarimo EAM, Bakari M, Sandström E, Kulane A. Recruitment using respondent driven sampling, risk behaviors assessment and willingness of young female sex workers (18-25 years) in Dar Es Salaam, Tanzania to participate in HIV vaccine trials. BMC Public Health. 2019 Nov 20;19(1):1537.
186. Merson MH, O'Malley J, Serwadda D, Apisuk C. The history and challenge of HIV prevention. Lancet. 2008 Aug 9;372(9637):475–88.
187. Broadhead RS, Heckathorn DD, Weakliem DL, Anthony DL, Madray H, Mills RJ, et al. Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention. Public Health Rep. 1998 Jun;113 Suppl 1:42–57.