†These authors equally contributed to this study as co-first authors.
‡These authors equally contributed to this study as co-corresponding authors.

THE ROYAL SOCIETY PUBLISHING

Modelling the geographical spread of HIV among MSM in Guangdong, China: a metapopulation model considering the impact of pre-exposure prophylaxis

Fengshi Jing[1,2,3,†], Yang Ye[3,†], Yi Zhou[4,5,†], Hanchu Zhou[3,6], Zhongzhi Xu[7], Ying Lu[2], Xiaoyu Tao[4], Shujuan Yang[8], Weibin Cheng[1,3], Junzhang Tian[1], Weiming Tang[1,2,‡] and Dan Wu[2,8,9,‡]

[1]Institute for Healthcare Artificial Intelligence, Guangdong Second Provincial General Hospital, Guangzhou 510317, People's Republic of China
[2]University of North Carolina Project-China, Guangzhou 510085, People's Republic of China
[3]School of Data Science, City University of Hong Kong, Hong Kong SAR, People's Republic of China
[4]Faculty of Medicine, Macau University of Science and Technology, Macau SAR, People's Republic of China
[5]Zhuhai Center for Diseases Control and Prevention, Zhuhai 519060, People's Republic of China
[6]School of Traffic and Transportation Engineering, Central South University, Changsha 410075, People's Republic of China
[7]The Hong Kong Jockey Club Centre for Suicide Research and Prevention, The University of Hong Kong, Hong Kong SAR, People's Republic of China
[8]West China School of Public Health, Sichuan University, Chengdu 610041, People's Republic of China
[9]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

FJ, 0000-0002-6747-6527; YY, 0000-0003-0710-1341; HZ, 0000-0002-6305-7907; ZX, 0000-0002-0514-0200; WT, 0000-0002-9026-707X; DW, 0000-0003-0415-5467

Men who have sex with men (MSM) make up the majority of new human immunodeficiency virus (HIV) diagnoses among young people in China. Understanding HIV transmission dynamics among the MSM population is, therefore, crucial for the control and prevention of HIV infections, especially for some newly reported genotypes of HIV. This study presents a metapopulation model considering the impact of pre-exposure prophylaxis (PrEP) to investigate the geographical spread of a hypothetically new genotype of HIV among MSM in Guangdong, China. We use multiple data sources to construct this model to characterize the behavioural dynamics underlying the spread of HIV within and between 21 prefecture-level cities (i.e. Guangzhou, Shenzhen, Foshan, etc.) in Guangdong province: the online social network via a gay social networking app, the offline human mobility network via the Baidu mobility website, and self-reported sexual behaviours among MSM. Results show that PrEP initiation exponentially delays the occurrence of the virus for the rest of the cities transmitted from the initial outbreak city; hubs on the movement network, such as Guangzhou, Shenzhen, and Foshan are at a higher risk of 'earliest' exposure to the new HIV genotype; most cities acquire the virus directly from the initial outbreak city while others acquire the virus from cities that are not initial outbreak locations and have relatively high betweenness centralities, such as Guangzhou, Shenzhen and Shantou. This study provides insights in predicting the geographical spread of a new genotype of HIV among an MSM population from different regions and assessing the importance of prefecture-level cities in the control and prevention of HIV in Guangdong province.

This article is part of the theme issue 'Data science approach to infectious disease surveillance'.

1. Introduction

Despite substantial progress in tackling the human immunodeficiency virus (HIV) epidemic, HIV continues to pose public health threats in China [1,2]. Specifically, HIV transmission among men who have sex with men (MSM) has increased markedly in recent years, making up the majority of new diagnoses among young people in China [3–5]. HIV mutates frequently. Some genotypes may become more infectious and virulent than the existing ones, which causes current antiretroviral treatment (ART) to be less effective and a larger number of people to be infected [6,7]. Increasing connectivity among cities has further complicated HIV transmission. Understanding transmission patterns of a new genotype of HIV among MSM, especially the geographical spread patterns, is important for the control and prevention of HIV infections. An example of the outbreak of a new genotype in recent years is HIV-1 CRF55_01B, which was first reported in 2012 and was later proven to have originated in MSM in Shenzhen, a city of Guangdong province in China [6,7]. This genotype has been causing thousands of HIV infection cases in the whole Guangdong province through an inter-city geographical transmission pattern from Shenzhen. A surveillance system to alert neighbouring cities about the emergence of a new HIV genotype may help us better control and prevent HIV transmission across cities. Therefore, the motivation of this study is to first construct such an alarming system (i.e. an inter-city probabilistic transmission network model) for cities by hypothesizing the emergence of a new HIV genotype in a certain city and to predict possible geographical spread using a metapopulation model.

Metapopulation models are widely used to investigate the transmission of infectious diseases, such as influenza, the 2003 SARS epidemic, and the COVID-19, in a large-scale spatial area, where disease transmission among spatially separated populations occurs via population movement [8–12]. Previous research has used metapopulation models to measure the effect of human mobility on HIV transmission, to evaluate the effectiveness of different intervention strategies, and to predict further transmission [13–17]. For example, Megan *et al.* employed the metapopulation approach to estimate the impact of migration on the spread of HIV based on

a population study of migrant, non-migrant men, and their rural partners, in South Africa [13]. They concluded that migration influences HIV spread by increasing likelihood of high-risk sexual behaviours in their destination cities. Xiao *et al.* assessed the HIV epidemic in mainland China based on a mobility network of diagnosed HIV/AIDS cases, which was constructed using differences in their current residences and registered residences in China surveillance systems [14]. Isdory *et al.* built a Susceptible-Infected-Removed (SIR) based metapopulation model to analyse the impact of human mobility on HIV transmission based on the mobile phone location data in Kenya [15]. They showed that mobility between regions played a relatively small role in HIV transmission in Kenya where HIV has already been endemic.

However, little has been done to examine the geographical spread of HIV among the MSM population under the context where HIV prevention measures are provided to at-risk populations, such as HIV pre-exposure prophylaxis (PrEP) in China [18]. Effective HIV prevention measures including sexuality education, condom use promotion, and PrEP can significantly lower risks for transmitting the HIV virus. We focus on the MSM population and the usage of PrEP in this model due to the high HIV prevalence among MSM in China and the fairly large number of Chinese MSM who are willing to take PrEP [19]. To investigate the geographical spread of a new HIV genotype among MSM, we construct a movement probabilistic network to infer the MSM movement patterns between cities in Guangdong, one of the major provinces with a high HIV incidence among MSM in China [20]. Based on this network, we propose a metapopulation model, considering the impact of PrEP and parameterized with sexual behaviour data among MSM in Guangdong, to mimic the HIV transmission patterns between and within cities. This study aims to predict the geographical spread of a new genotype of HIV among MSM populations from different cities, and to assess the importance of each city in the control and prevention of HIV infections through simulations considering the impact of PrEP. Note that our approach can be adopted to predict geographical spread at a larger scale (such as transmission across provinces and countries). In this paper we used inter-city geographical spread in Guangdong province to illustrate the model building process.

2. Inter-city movement probabilistic network

Because (i) MSM with follower–followee relationships in online social media may have real-life contacts [21], and (ii) MSM population may have similar mobility patterns to other populations, we constructed an inter-city movement probabilistic network among MSM which integrates the multi-source data of online social networks and the offline mobility network to capture the MSM movement patterns between 21 prefecture-level cities (we use 'cities' thereafter for simplicity) in Guangdong. The offline mobility network is constructed using data from Baiduqianxi [22], a public website providing a real-time migration index based on the number of inbound and outbound events by rail, air and road traffic across mainland China for the entire migrant population (not just MSM migrants). Here, a node represents a city. An edge between nodes represents the average number of individuals travelling between two cities. We specify the number of MSM travelling between cities on a pro rata basis: denote the total number of migrants travelling between city i and city j per time step as M_{ij}, then the number of MSM migrants travelling between city i and city j per time step is δM_{ij}, where δ denotes the percentage of MSM population among the whole population for all cities (see electronic supplementary material for details). The online social network is constructed using data from Blued, the biggest gay men's dating app in China [23]. We randomly select 2011 users located in Guangdong province to infer the general online inter-city social network. Here, a node also represents a city. An edge between two cities represents the number of follower–followee relationships between two cities. Figure 1 illustrates the weighted degree distribution of the offline mobility network and the online social network. We can observe the long-tailed (power-low) degree distribution on both networks, and the offline mobility network has a longer tail than the online social network. We illustrate the construction of the movement probabilistic network in figure 2. We denote X_{ij}, Y_{ij}, W_{ij} as the adjacency matrix of the online social network, the offline mobility network, and the movement
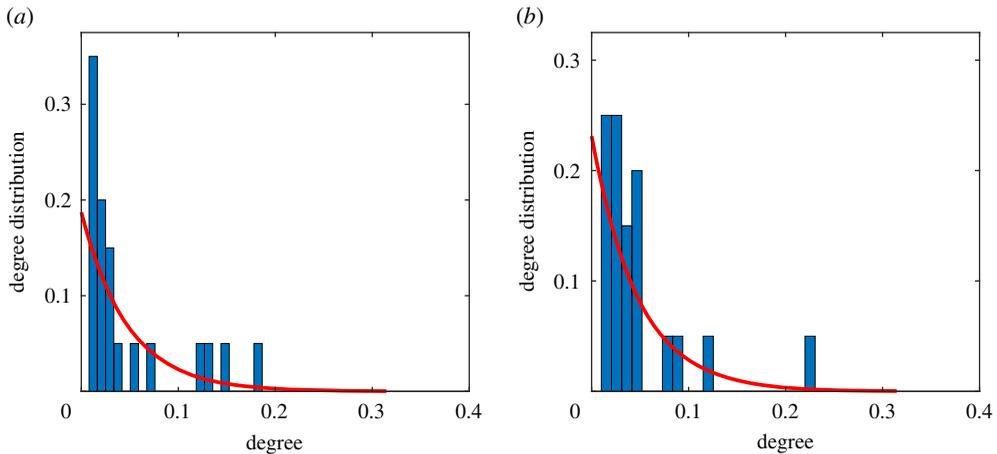
Figure 1. Weighted degree distribution of (*a*) the offline mobility network and (*b*) the online social network. For clear illustration, the weighted degree of each city is rescaled by dividing through the sum of the weighted degree of all cities. (Online version in colour.)

probabilistic network. Then, we use the following weighted linear combination to infer the actual movement patterns among MSM.

$$W_{ij} = \alpha \frac{X_{ij}}{X_i} + (1 - \alpha) \frac{Y_{ij}}{Y_i}, \tag{2.1}$$

where $X_i = \sum_j X_{ij}$, $Y_j = \sum_i Y_{ij}$, and α is the relative weight of the online social network in deciding the actual movement. Here, W_{ij} indicates the movement preference for MSM in city i [24]. Specifically, if one of MSM in city i moves, he goes to city j with a probability W_{ij}. $W_{ii} = 0$ and $\sum_j W_{ij} = 1$. According to equation (2.1), we assume such movement preference is determined by the weight of the interaction on the online social network and the offline mobility network. Note that, although the online social network and the offline mobility network are both symmetric and undirected networks, i.e. $X_{ij} = X_{ji}$ and $Y_{ij} = Y_{ji}$, the movement probabilistic network is an asymmetric and directed network, i.e. $W_{ij} \neq W_{ji}$. Owing to limited data, we cannot obtain a specific value of α that describes the accurate combination of the online social network and the offline mobility network. We will present and compare the geographical spread patterns under different combinations (i.e. different values of α) in this study.

3. The Meta-Spudtr model

To investigate transmission patterns of a new HIV genotype within and between cities, we propose Meta-Spudtr, a metapopulation model where the whole MSM population in Guangdong is divided by a set of subpopulations connected by moving individuals (figure 3). To be specific, we follow the prefecture-level divisions, with each subpopulation corresponding to all MSM in a city. We describe the disease dynamics within each city based on a deterministic dynamic compartmental model in [25]. Individuals in each city are divided into six classes: susceptible, on PrEP, living with undiagnosed HIV, living with diagnosed HIV, on HIV ART, and removed. We illustrate the transitions between states (e.g. 'Susceptible', 'On PrEP', etc.) in figure 3.

Denote $S_i(t)$, $P_i(t)$, $U_i(t)$, $D_i(t)$, $T_i(t)$ and $R_i(t)$ as the number of susceptible, on pre-exposure PrEP, living with undiagnosed HIV, living with diagnosed HIV, on HIV ART, and removed individuals at time t for city i, respectively. $N_i = S_i(t) + P_i(t) + U_i(t) + D_i(t) + T_i(t) + R_i(t)$ is the total number of individuals in city i. Susceptible individuals will initiate PrEP at a rate σ and we assume that all individuals will not stop PrEP once starting. Susceptible and on PrEP individuals may acquire HIV through sexual activities with individuals living with undiagnosed HIV and
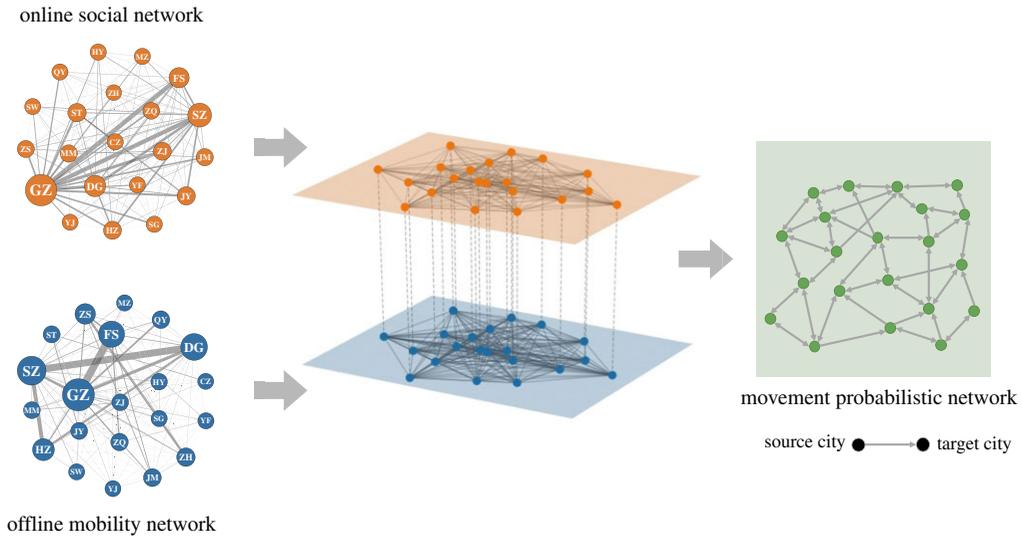
Figure 2. Construction of the movement probabilistic network. The panels in the left column show the visualizations of the online social network (top one; obtained from BlueD) and the offline mobility network (bottom one; obtained from Baiduqianxi). A node represents a city in Guangdong, China. Node labels represent the abbreviations of city names. A detailed mapping of city names and abbreviations is presented in the electronic supplementary material. An edge on online social network represents the number of follower–followee relationships between two cities. An edge on the offline mobility network represents the number of individuals travelling between two cities. The movement probabilistic network on the right column is constructed by integrating the information on the online social network and the offline mobility network (the middle panel); see equation (2.1). A directed edge on the movement probabilistic network represents the probability of individuals moving from the source city to the target city. (Online version in colour.)
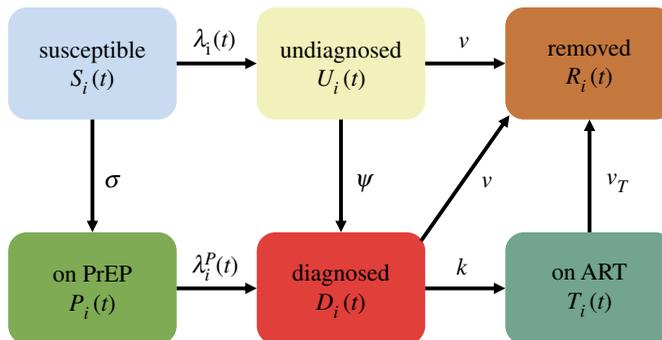

Figure 3. Illustration of the transitions between states within cities. Definitions for parameters in figure 3 are provided in the electronic supplementary material, equations (3.1) and (3.2). (Online version in colour.)

diagnosed HIV. We hypothesize that the risks for transmitting HIV from people living with HIV but on regular ART treatment to susceptible individuals are minimal [26]. We also make an assumption that once diagnosed as HIV-positive, individuals will decrease the number of sexual contacts and use condoms in sexual activities. Denote the rates at which susceptible and on PrEP individuals acquire HIV are $\lambda_i(t)$ and $\lambda_i^P(t)$, respectively. For susceptible individuals,

$$\lambda_i(t) = C\beta_S \left[\frac{U_i(t)}{N_i} + (1 - \theta)(1 - \eta_C)\frac{D_i(t)}{N_i} \right], \tag{3.1}$$

where C is the average number of contacts per person per time, β_S is the transmission probability per anal sex act, θ is the relative decrease change in the number of contacts for individuals living with diagnosed HIV compared to those living with undiagnosed HIV, η_C is the condom efficacy among MSM. Similarly, for individuals on PrEP,

$$\lambda_i^P(t) = C\beta_P \left[\frac{U_i(t)}{N_i} + (1-\theta)(1-\eta_C)\frac{D_i(t)}{N_i} \right], \tag{3.2}$$

where $\beta_P = \eta_P \beta_S$ is the transmission probability per anal sex act with condom use. η_P is the efficacy of PrEP. After being infected with HIV, on PrEP individuals will immediately become living with diagnosed HIV, while susceptible ones will first become undiagnosed and then transit to diagnosed at a rate ψ. The mortality rates for individuals living with undiagnosed HIV and diagnosed HIV are the same, denoted by v. Individuals living with diagnosed HIV will initiate ART at a rate k and we assume that no one will drop out once starting. The mortality rate for individuals on ART is denoted by v_T. In addition to transmission dynamics within each city, HIV can spread between cities through individual movements. Denote $\gamma = \sum_{i,j} Y_{ij} / \sum_i N_i$ as the average movement rate, i.e. each MSM decides whether or not to move with probability γ or remain in current city with probability $1 - \gamma$. If he moves, the choice of destination is decided by W_{ij} obtained from equation (2.1). Then, we can obtain the differential equations to describe HIV transmission dynamics within and between cities as follows:

$$\partial_t S_i(t) = -[\lambda_i(t) + \sigma]S_i(t) + \gamma N_i \sum_{j\neq i} W_{ij}\left[\frac{S_j(t)}{N_j} - \frac{S_i(t)}{N_i} \right],$$

$$\partial_t P_i(t) = \sigma S_i(t) - \lambda_i^P(t)P_i(t) + \gamma N_i \sum_{j\neq i} W_{ij}\left[\frac{P_j(t)}{N_j} - \frac{P_i(t)}{N_i} \right],$$

$$\partial_t U_i(t) = \lambda_i(t)S_i(t) - (\psi + v)U_i(t) + \gamma N_i \sum_{j\neq i} W_{ij}\left[\frac{U_j(t)}{N_j} - \frac{U_i(t)}{N_i} \right],$$

$$\partial_t D_i(t) = \psi U_i(t) + \lambda_i^P(t)P_i(t) - (k+v)D_i(t) + \gamma N_i \sum_{j\neq i} W_{ij}\left[\frac{D_j(t)}{N_j} - \frac{D_i(t)}{N_i} \right], \tag{3.3}$$

$$\partial_t T_i(t) = kD_i(t) - v_T T_i(t) + \gamma N_i \sum_{j\neq i} W_{ij}\left[\frac{T_j(t)}{N_j} - \frac{T_i(t)}{N_i} \right]$$

and

$$\partial_t R_i(t) = v[U_i(t) + D_i(t)] + v_T T_i(t) + \gamma N_i \sum_{j\neq i} W_{ij}\left[\frac{R_j(t)}{N_j} - \frac{R_i(t)}{N_i} \right].$$

The model works in a monthly time step. Note that (a) there is no entry into or departure from the whole MSM population; (b) the model is restricted to the transmission of the new genotype of HIV, thus, infections with other genotypes are not considered; (c) we only consider HIV transmission among MSM (i.e. via anal sex) in this model for simplicity. The values of epidemiological parameters in equation (3.3) are estimated mainly based on previous sexual behavioural surveys and reports of MSM in Guangdong, as all cities in Guangdong province are selected for modelling. Corresponding references are provided in the electronic supplementary material.

4. Simulations and results

(a) General transmission patterns

Figure 4 presents the total numbers of MSM living with HIV (i.e. undiagnosed, diagnosed but not on ART, and diagnosed and on ART) and the distribution of infected cases in the five most
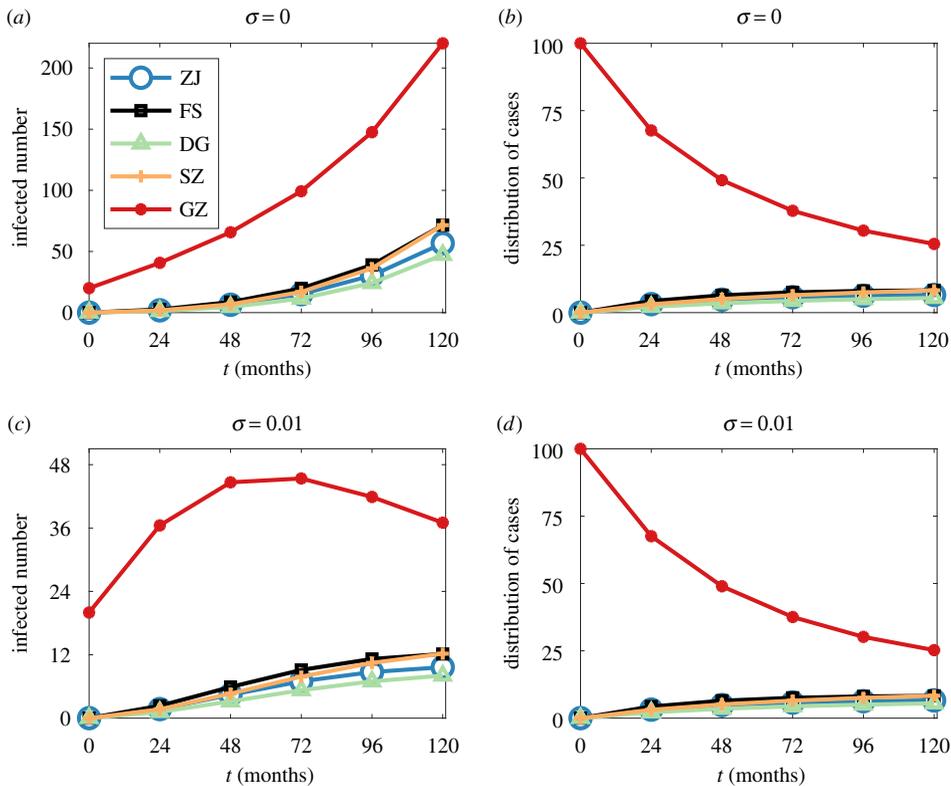
Figure 4. Total numbers of infected MSM (i.e. MSM living with HIV) in GZ, SZ, DG, FS and ZJ with (*a*) $\sigma = 0$ and (*c*) $\sigma = 0.01$. Distribution of infected MSM in GZ, SZ, DG, FS and ZJ with (*b*) $\sigma = 0$ and (*d*) $\sigma = 0.01$. $\alpha = 0.5$, Guangzhou is initially infected with 20 MSM at $t = 0$. GZ, Guangzhou; SZ, Shenzhen; DG, Dongguan; FS, Foshan; ZJ, Zhanjiang. (Online version in colour.)

populous cities (i.e. Guangzhou, Shenzhen, Dongguan, Foshan and Zhanjiang) during the first 10 years with 20 MSM initially infected in Guangzhou. Although all five cities report infections within 2 years, most new infections still occur in the initial outbreak city, Guangzhou, within the first 10 years. The growth rate of infected cases for each city is partially relevant to the movement probability from the initial outbreak city to other cities: the movement probability from Guangzhou to Foshan (0.23) is greater than that to Dongguan (0.11), thus, the growth rate at Foshan is higher than that at Dongguan. Of note, the movement probabilities from Guangzhou to Dongguan and Shenzhen are almost the same, but we observe a slightly higher growth rate in Shenzhen than that in Dongguan. Besides, PrEP initiation has a significant effect in preventing infections: the number of infected cases in Guangzhou decreases gradually from the fifth year when σ is 0.01, while it keeps growing when σ is 0. Although the numbers of infected cases in other cities still grow when σ is 0.01, we can observe a much lower increase compared to the scenario when σ is 0 and the increase rate is getting slower. Details about the difference in the time series of infected cases are presented in the electronic supplementary material.

(b) Impact of PrEP

To quantitatively explain the impact of PrEP and the dynamics underlying the geographical spread of HIV, we provide a simpler framework to capture the dynamics in equation (3.3). Essentially, if the PrEP initiation rate $\sigma = 0$, the Meta-Spudtr model is approximately an SIR based metapopulation model. Previous studies have proposed a framework that can reduce the complex spatiotemporal patterns of such disease transmission by a simple propagation pattern.

Specifically, the disease arrival time can be predicted by a probabilistically motivated effective distance [9,27]. Following this literature, we define the effective distance from city i to a connected city j as

$$d_{ij} = (1 - \log W_{ij}) \geq 1. \tag{4.1}$$

Note that, although the definition of W_{ij} is a bit different from that in [9], they both quantify the fraction of individuals with destination city j emanating from city i. Equation (4.1) is consistent with the concept in [9] that a small W_{ij} is equivalent to a large effective distance, and vice versa. Since $W_{ij} \neq W_{ji}$, then $d_{ij} \neq d_{ji}$. Then, we can obtain the length $\Lambda(\Gamma)$ (based on the effective distance) of a directed path $\Gamma = \{i_1, \ldots, i_L\}$ as

$$\Lambda(\Gamma) = \sum_{m=1}^{m=L-1} d_{i_m i_{m+1}}, \tag{4.2}$$

Then, we define the effective distance from city i to an arbitrary city j as the shortest length among all directed paths from city i to city j, i.e.

$$D_{ij} = \min_{\Gamma} \Lambda(\Gamma). \tag{4.3}$$

Assuming the initial outbreak city of a specific genotype of HIV is city i, this genotype arrives in an arbitrary city j at time T_{ij}^a, the framework in [9] suggests a linear relationship between the disease arrival time T_{ij}^a and the effective distance D_{ij}, i.e.

$$T_{ij}^a \propto D_{ij}. \tag{4.4}$$

The arrival time is defined as the date of the first infected individual (living with undiagnosed or diagnosed HIV). Figure 5a demonstrates the relationship between T^a and D with the initial outbreak in Zhuhai, $\alpha = 0.7$, and $\sigma = 0$. We observe a high value of R^2, indicating that the HIV transmission patterns can be well characterized by equation (4.4) with no one taking PrEP. However, as σ grows, T^a correlates weakly with D. Figure 5b demonstrates the relationship between T^a and D when $\sigma = 0.03$, other parameters are the same as that in figure 5a. We observe that R^2 decreases greatly, dropping to 0.61 in figure 5b from 0.87 in figure 5a. This indicates that, since the framework in [9] did not consider the impact of prevention tools, equation (4.4) cannot perfectly characterize the underlying dynamics of HIV transmission as σ grows.

Here, we propose a new framework that can better capture the geographical spread patterns of HIV among MSM considering the impact of PrEP. We modified the framework in [9] to characterize the relationship between the disease arrival time T_{ij}^a and the effective distance D_{ij} as

$$\log T_{ij}^a \propto D_{ij}. \tag{4.5}$$

Figure 5 demonstrates the comparison of equations (4.4) and (4.5) in describing the relationship between the arrival time T^a and the effective distance D with different initial outbreak cities, different combinations of online social network and offline mobility network, and different PrEP initiation rates. We can observe in all subfigures that, due to the impact of PrEP ($\sigma > 0$), effective distance D has a stronger correlation with the natural logarithm of the arrival time $\log T^a$ than the arrival time T^a itself. More similar results for a wide range of σ are presented in the electronic supplementary material. These indicate that (a) given an initial outbreak city, the order in which each city is infected by HIV transmission depends only on the underlying movement probabilistic network regardless of the PrEP initiation rate σ; (b) PrEP initiation exponentially delays the occurrence of a specific genotype of HIV for all cities.

(c) 'Earliest' exposure risk

According to the linear relationship between $\log T^a$ and D, virus occurrence delays are much shorter in cities with shorter effective distances from the initial outbreak city. Thus, cities with shorter distances from other cities remain at high risk of immediate exposure to the virus,
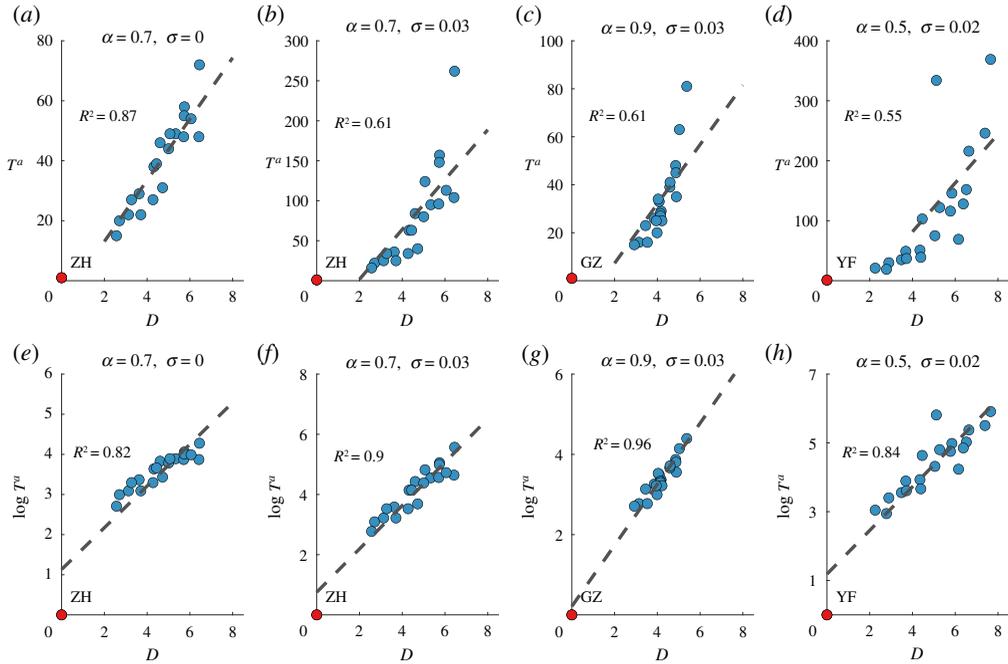
Figure 5. The relationship between the HIV arrival time T^a and the effective distance D when initial outbreak city is Zhuhai, $\alpha = 0.7$, $\sigma = 0$ in (*a*) and (*e*); initial outbreak city is Zhuhai, $\alpha = 0.7$, $\sigma = 0.03$ in (*b*) and (*f*); initial outbreak city is Guangzhou, $\alpha = 0.9$, $\sigma = 0.03$ in (*c*) and (*g*); initial outbreak city is Yunfu, $\alpha = 0.5$, $\sigma = 0.02$ in (*d*) and (*h*). The number of initially infected MSM is 20. GZ, Guangzhou; ZH, Zhuhai and YF, Yunfu. (Online version in colour.)

even with the impact of PrEP. An earlier occurrence of the virus indicates a shorter time for preparedness against the virus. Here, we define the 'earliest' exposure risk for each city by the probability of being the first infected city except the initial outbreak city. Given a movement probabilistic network (fixed α), the 'earliest' infection risk for city i EIR_i is denoted as

$$\mathrm{EIR}_i = \frac{\sum_j I(D_{ji} = \min_{j \neq k} D_{jk})}{n}, \tag{4.6}$$

where n is the number of cities, $I(\cdot)$ is the indicator function. $\sum_i \mathrm{EIR}_i = 1$. Figure 6 illustrates the 'earliest' exposure risk for each city with respect to different values of α. We can observe that Guangzhou, Shenzhen, Foshan and Shantou will always face the risk of 'earliest' exposure regardless of the value of α. The differences in the 'earliest' exposure risks for cities can also be well explained by their positions on the movement network. Here, we plot the relationship between EIR and the weighted degree for each city with $\alpha = 0$ and $\alpha = 1$, respectively, in figure 7. We can observe in both subfigures that hub cities usually face higher 'earliest' exposure risk. As α grows, fewer cities have large degrees while Guangzhou becomes a larger hub, thus, the 'earliest' exposure risk has been transferred to Guangzhou from other cities. Compared to figure 6*a*, we can observe that fewer cities face the 'earliest' exposure risk while Guangzhou faces a much higher 'earliest' exposure risk in figure 6*c*.

(d) Most possible transmission path

By converting weights of edges on the movement probabilistic network to the effective distance based on equation (4.1), we can obtain an effective distance-based movement network among MSM. The shortest path tree generating from the effective distance-based movement network indicates the most possible transmission path from the initial outbreak city to a random
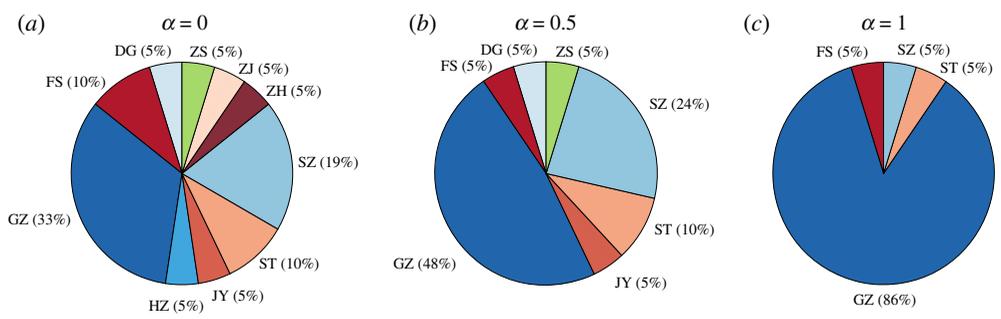
Figure 6. The 'earliest' exposure risk for each city with respect to (a) $\alpha = 0$, (b) $\alpha = 0.5$, and (c) $\alpha = 1$. Cities with $EIR_i = 0$ are not presented in the figure. GZ, Guangzhou; ZH, Zhuhai; HZ, Huizhou; JY, Jieyang; ST, Shantou; SZ, Shenzhen; ZH, Zhuhai; ZJ, Zhanjiang; ZS, Zhongshan; DG, Dongguan; FS, Foshan. (Online version in colour.)

Figure 7. Relationship between EIR and the weighted degree for each city with (a) $\alpha = 0$ and (b) $\alpha = 1$. The weighted degree for each city is rescaled by dividing through the sum of the weighted degree of all cities. GZ, Guangzhou; ZH, Zhuhai; HZ, Huizhou; JY, Jieyang; ST, Shantou; SZ, Shenzhen; ZH, Zhuhai; ZJ, Zhanjiang; ZS, Zhongshan; DG, Dongguan; FS, Foshan. (Online version in colour.)
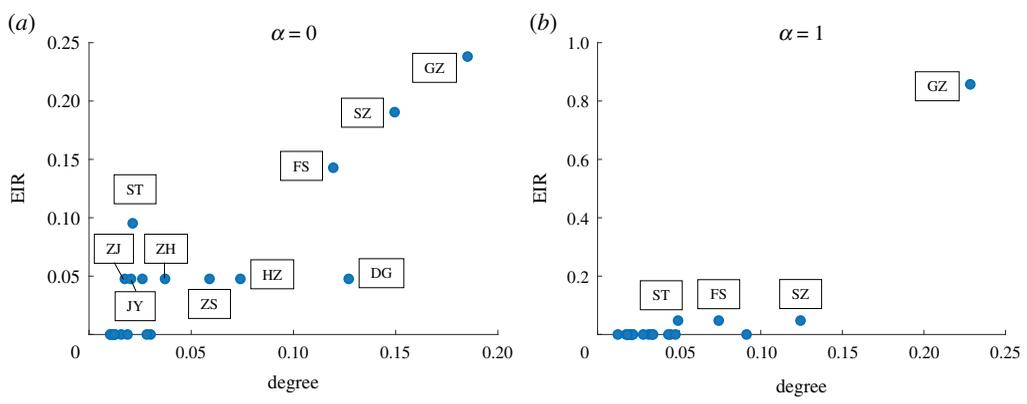
city. Figure 8 illustrates the shortest path tree with different values of α and different initial outbreak cities. Here, we only present the results for $\alpha \in \{0, 0.5, 1\}$ and initial outbreak in {ST (Shantou), HY (Heyuan)}. Results for other scenarios are consistent with the results in figure 8. We can observe that most cities acquire the virus directly from the initial outbreak city, while some cities acquire the virus from non-initial outbreak cities like Guangzhou, Shenzhen and Shantou. Figure 9 illustrates the betweenness centralities [28] for these three cities on the effective distance-based movement network with different values of α. As α increases, all betweenness centralities for Guangzhou, Shenzhen and Shantou increase. Besides, Guangzhou always has a higher betweenness centrality than Shenzhen and Shantou. These results quantitatively demonstrate the importance of each city in facilitating HIV transmission between cities. Containment of the outbreak in cities with high betweenness centralities slows the contagion of the virus, because several cities can only reach each other via longer paths. It is not surprising that Guangzhou, the capital of Guangdong province and Shenzhen, the first special economic zone in China, play a huge role in bridging cities and facilitating transmission. However, Shantou, as a medium sized city in Guangdong province, should also not be neglected for control and prevention measures.

Figure 8. The shortest path tree for (a) $\alpha = 0$, initial outbreak in ST, (b) $\alpha = 0.5$, initial outbreak in ST, (c) $\alpha = 1$, initial outbreak in ST, (d) $\alpha = 0$, initial outbreak in HY, (e) $\alpha = 0.5$, initial outbreak in HY, and (f) $\alpha = 1$, initial outbreak in HY. ST, Shantou; HY, Heyuan. (Online version in colour.)

Figure 9. Betweenness centrality for GZ (Guangzhou), SZ (Shenzhen) and ST (Shantou) on the effective distance-based movement network. Other cities with betweenness centralities of zero are not presented here. (Online version in colour.)

5. Conclusion

In this study, we have developed the Meta-Spudtr, a metapopulation model considering the impact of PrEP, to investigate the transmission patterns of a specific genotype of HIV among MSM in Guangdong, China. Our Meta-Spudtr model works based on an inter-city movement probabilistic network among MSM, which integrates the offline mobility network and the online social network among MSM. Based on previous studies [9], we propose a new framework that can

quantitatively capture the geographical spread patterns of HIV under the circumstances when PrEP services are provided to MSM.

We find that (a) although the virus arrives in some cities 2 years after the initial outbreak, the majority of new infections remain in the initial outbreak city in the first 10 years; (b) PrEP initiation will not change the order in which cities will be infected; (c) but PrEP initiation exponentially delays the occurrence of a new HIV genotype in all rest cities transmitted from the outbreak city; (d) due to the hub-effect, Guangzhou, Shenzhen, Foshan and Shantou will always face the risk of 'earliest' exposure in all combinations of the offline mobility network and the online social network; (e) most cities acquire the virus directly from the initial outbreak city, while some cities acquire the virus from cities that are not initial outbreak locations but have relatively high betweenness centralities on the effective distance-based movement network, such as Guangzhou, Shenzhen and Shantou. Our study extends previous studies [9] to predict the arrival times of a new HIV genotype considering the impact of PrEP. It also assesses the exposure risk and measures the importance in facilitating transmission for each city in Guangdong. This framework can be easily extended to investigate the large-scale transmission patterns of other infectious diseases with the impact of intervention tools by quantifying the underlying movement probabilistic network and the related epidemiological parameters. Our model has limitations. First, we do not consider the heterogeneity in PrEP initiation rates, HIV testing rates, ART initiation rates among cities, and the sexual behaviour preferences, contact patterns among MSM. Incorporating such heterogeneity will be interesting future work. Second, due to limited access to administrative data, we combine online social networking apps and public website documenting human mobility for estimation, and the value of α cannot be accurately determined. Further studies should be conducted for calibrations of the current model in real-world applications.

(All identifiable personal information was removed for privacy protection.)

References

1. UNAIDS. 2019 *2019 China AIDS response progress report*. See www.unaids.org/sites/default/files/country/documents/CHN.
2. Zheng S. 2018 The growing threat of China's HIV epidemic. *Lancet Public Health* **3**, e311. (doi:10.1016/S2468-2667(18)30098-7)
3. Dong MJ, Peng B, Liu ZF, Ye Qn, Liu H, Lu XL, Zhang B, Chen JJ. 2019 The prevalence of HIV among MSM in china: a large-scale systematic analysis. *BMC Infect. Dis.* **19**, 1–20. (doi:10.1186/s12879-018-3567-x)
4. Wu Z, Chen J, Scott SR, McGoogan JM. 2019 History of the HIV epidemic in China. *Curr. HIV/AIDS Rep.* **16**, 458–466. (doi:10.1007/s11904-019-00471-4)

5. Booton RD *et al.* 2021 Modelling the impact of an HIV testing intervention on HIV transmission among men who have sex with men in China. *HIV Med.* **22**, 467–477. (doi:10.1111/hiv.13063)

6. Zhao J. *et al.* 2016 The dynamics of the HIV epidemic among men who have sex with men (MSM) from 2005 to 2012 in Shenzhen, China. *Sci. Rep.* **6**, 1–8. (doi:10.1038/s41598-016-0001-8)

7. Yu G *et al.* 2020 Genetic diversity and drug resistance of HIV-1 crf55_01b in Guangdong, China. *Curr. HIV Res.* **18**, 210–218. (doi:10.2174/1570162X18666200415140652)

8. Pei S, Kandula S, Yang W, Shaman J. 2018 Forecasting the spatial transmission of influenza in the United States. *Proc. Natl Acad. Sci. USA* **115**, 2752–2757. (doi:10.1073/pnas.1708856115)

9. Brockmann D, Helbing D. 2013 The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342. (doi:10.1126/science.1245200)

10. Ruktanonchai NW *et al.* 2020 Assessing the impact of coordinated Covid-19 exit strategies across Europe. *Science* **369**, 1465–1470. (doi:10.1126/science.abc5096)

11. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. 2021 Mobility network models of Covid-19 explain inequities and inform reopening. *Nature* **589**, 82–87. (doi:10.1038/s41586-020-2923-3)

12. Liao H, Zhang L, Marley G, Tang W. 2020 Differentiating Covid-19 response strategies. *Innovation* **1**, 100003. (doi:10.1016/j.xinn.2020.04.003)

13. Coffee M, Lurie MN, Garnett GP. 2007 Modelling the impact of migration on the HIV epidemic in South Africa. *Aids* **21**, 343–350. (doi:10.1097/QAD.0b013e328011dac9)

14. Xiao Y, Tang S, Zhou Y, Smith RJ, Wu J, Wang N. 2013 Predicting the HIV/Aids epidemic and measuring the effect of mobility in mainland China. *J. Theor. Biol.* **317**, 271–285. (doi:10.1016/j.jtbi.2012.09.037)

15. Isdory A, Mureithi EW, Sumpter DJ. 2015 The impact of human mobility on HIV transmission in kenya. *PLoS ONE* **10**, e0142805. (doi:10.1371/journal.pone.0142805)

16. Smith RJ, Li J, Gordon R, Heffernan JM. 2009 Can we spend our way out of the AIDS epidemic? A world halting AIDS model. *BMC Public Health* **9**, 1–17. (doi:10.1186/1471-2458-9-S1-S1)

17. Sun X, Xiao Y, Peng Z, Wang N. 2018 Frequent implementation of interventions may increase HIV infections among MSM in China. *Sci. Rep.* **8**, 1–11. (doi:10.1038/s41598-017-18743-7)

18. Anderson PL *et al.* 2012 Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. *Sci. Transl. Med.* **4**, 151ra125. (doi:10.1126/scitranslmed.3004006)

19. Xu J, Tang W, Zhang F, Shang H. 2020 Prep in China: choices are ahead. *Lancet HIV* **7**, e155–e157. (doi:10.1016/S2352-3018(19)30293-0)

20. Huang W, Wang Y, Lu H, Wu D, Pan SW, Tucker JD, Tang W. 2020 High HIV incidence among men who have sex with men in 8 Chinese cities: results from a trial. In *Open forum infectious diseases* (ed. PE Sax), vol. 7, p. ofaa147. Oxford University Press US.

21. Young SD, Nianogo RA, Chiu CJ, Menacho L, Galea J. 2016 Substance use and sexual risk behaviors among peruvian MSM social media users. *AIDS Care* **28**, 112–118. (doi:10.1080/09540121.2015.1069789)

22. Baidu. 2020 *Baidu Qianxi*. See https://qianxi.baidu.com/.

23. Blued. See https://blued.cn/.

24. Soriano-Paños D, Lotero L, Arenas A, Gómez-Gardeñes J. 2018 Spreading processes in multiplex metapopulations containing different mobility networks. *Phys. Rev. X* **8**, 031039. (doi:10.1103/PhysRevX.8.031039)

25. Wang L *et al.* 2021 Mathematical modelling of the influence of serosorting on the population-level HIV transmission impact of pre-exposure prophylaxis. *AIDS* **35**, 1113. (doi:10.1097/QAD.0000000000002826)

26. Rodger AJ *et al.* 2016 Sexual activity without condoms and risk of hiv transmission in serodifferent couples when the HIV-positive partner is using suppressive antiretroviral therapy. *Jama* **316**, 171–181. (doi:10.1001/jama.2016.5148)

27. Wang L, Wu JT. 2018 Characterizing the dynamics underlying global spread of epidemics. *Nat. Commun.* **9**, 1–11. (doi:10.1038/s41467-017-02088-w)

28. Borgatti SP. 1995 Centrality and aids. *Connections* **18**, 112–114.