SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

Reassessing geographic bottlenecks in a respondent-driven sampling based multicity study in Brazil

Reevaluación de cuellos de botella geográficos en un estudio multiciudad basado en el método *respondent-driven sampling* en Brasil

Naide Teodósio Valois-Santos[1], Roberta Pereira Niquini[2], Sandro Sperandei[3], Leonardo Soares Bastos[4], Neilane Bertoni[5], Ana Maria de Brito[6], Francisco Inácio Bastos[7]

[1]***Corresponding author**: Public health physician. PhD in Collective Health. Associate Researcher, Instituto Aggeu Magalhães, Fundação Oswaldo Cruz Pernambuco, Brazil.*

[2]*Statistician. PhD in Epidemiology in Public Health. Professor, Instituto Federal de Educação, Ciência e Tecnologia, Rio de Janeiro, Brazil. Collaborating Researcher, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.*

[3]*Statistician. PhD in Computational Biology and Information Systems. PhD in Epidemiology, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.*

[4]*Statistician. PhD in Statistics. Associate Researcher in Public Health, Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil. Research Fellow, Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, England.*

[5]*Statistician. PhD in Epidemiology in Public Health. Biostatistician, Instituto Nacional do Câncer, Ministério da Saúde do Brasil. Collaborating Researcher, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.*

[6]*Public health physician. Tenured Researcher in Public Health, Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Pernambuco, Brazil.*

[7]*Public health physician. Tenured Researcher in Public Health, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.*

ABSTRACT This study analyzes the spatial dynamics of drug users' recruitment chains in the context of a respondent-driven sampling (RDS) study in the city of Recife, Brazil. The purpose is to understand the geographic bottlenecks, influenced by social geography, which have been a major challenge for RDS-based studies. Temporo-spatial analysis was used. Sequential maps depicted the dynamics of the recruiting process, considering neighborhood of residence and/or places of drug use. Poisson regression was fitted to model the recruiting rate by neighborhood of residence and/or places of drug use, and the different neighborhoods' demographics. The distance between neighborhood of residence and/or places of drug use and the assessment center was negatively associated with recruitment. There was a positive association between the proportion of the population living in informal settings and the recruiting rate per neighborhood of residence and/or places of drug use. Recruitment chains depend on the social geography and demographics of the population. Studies should incorporate seeds from as many neighborhoods as possible, and more than one assessment center should be utilized.

KEY WORDS Drug Users; Vulnerable Populations; Social Networks; Geographic Information Systems; Brazil.

RESUMEN Se analiza la dinámica espacial de las cadenas de reclutamiento de consumidores de drogas en un estudio *respondent-driven sampling* (RDS) en la ciudad de Recife, Brasil. El propósito es comprender los cuellos de botella geográficos, influenciados por la geografía social, que han sido un gran desafío para los estudios basados en RDS. Se utilizó el análisis espacio-temporal. La dinámica del proceso de reclutamiento se presenta en mapas, teniendo en cuenta el barrio de residencia y/o los sitios de consumo de drogas. La regresión de Poisson se ajustó para modelar la tasa de reclutamiento por barrio de residencia y/o sitios de consumo de drogas y las características demográficas de los barrios. La distancia entre el barrio de residencia y/o los sitios de consumo de drogas y el centro de evaluación se asoció negativamente con el reclutamiento. Hubo una asociación positiva entre la proporción de la población que vive en entornos informales y la tasa de reclutamiento por barrio de residencia y/o sitios de consumo de drogas. Las cadenas de reclutamiento dependen de la geografía social y demográfica de la población. Los estudios deben incorporar semillas de reclutamiento de tantos barrios como sea posible, así como más de un centro de evaluación.

PALABRAS CLAVES Consumidores de Drogas; Poblaciones Vulnerables; Red Social; Sistemas de Información Geográfica; Brasil.

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

BACKGROUND

People who use drugs, whose habits and behaviors are often associated with stigma, marginalization, besides being criminalized in several contexts, are defined as hard-to-reach or hidden populations. The magnitude, characteristics and boundaries of these populations remain to a great extent unknown, making an *a priori* specification of a sampling frame impossible. However, given the need to collect reliable and valid data on such populations in order to formulate and implement effective health interventions, alternative sampling methods have been proposed and implemented, such as Respondent Driven Sampling (RDS).[1,2,3,4,5]

RDS assumes that the members of hidden populations are the best recruiters of their own peers.[1] However, the process of mobilizing peers as recruiters involves, in an explicit or implicit way, the challenge of dealing with the underlying social structure. Social connections are established and reshaped over time not as a random process, but rather as one based on similarity or differential interaction rooted in characteristics which may include socioeconomic, demographic, cultural factors, among others.[2]

As recently discussed by Khabbazian et al.,[6] "even under the ideal sampling assumptions, the performance of RDS is restricted by the underlying social network: if the network is divided into weakly connected communities, then RDS is likely to oversample one of these communities."

The obvious consequence of oversampling a given subset among all potential communities/segments of interest, given the constraints imposed by the real-life conditions of operation of any empirical study,[7] is that other segments are likely to be undersampled or simply ignored. No matter what *ad hoc* improvements are implemented without violating the protocol and compromising the validity of the study (such as calibrating the distribution of study invitation "cupons" or recruiting additional seeds), bottlenecks will inevitably impose biases and caveats, the worst of them the possibility of bypassing a given targeted segment. Since segments do not tend to be homogeneous, the complex combination of over or undersampling may affect key variables such as infection rates for communicable diseases, which could either inflate or deflate such figures.

In a former study, we profited from a multicity study using RDS in 10 different Brazilian cities,[8] addressing one specific setting, Rio de Janeiro. The recruitment process was deeply affected by what was called by us "structural bottlenecks," that is, the fact that different parts of Rio's territory where our population of interest congregates (in our case, people who misuse illicit drugs, frequently engaged in the drug retail market) are "unbridgeable." By "unbridgeable" we mean here territories where factions engaged in feuds and conflicts preclude the potential circulation of study invitation cupons between different segments of social networks. People may know and sometimes interact with others, but such interactions are risky and/or strongly discouraged by faction leaders and their "watchers". As frequently stated by our interviewees, crossing lines could translate into death or serious threats to themselves or their families.

Several studies have been published after our own study, but to the best of our knowledge none targeted contexts which may resemble Brazilian drug scenes which are plagued by structural violence.[9] Other geographical bottlenecks have been identified and debated over time, such as those respecting Uganda villages,[10] rural areas of the US[11] or networks made up of both legal and illegal immigrants in the US.[12] Notwithstanding, none of these localities resemble Brazilian drug scenes, which may be the most violent worldwide, and where a sizeable fraction of over 60,000 registered violent deaths (in addition to non-registered casualties) take place every single year, all over the country, but especially in certain hotspots.[13]

In order to assess whether the patterns observed in Rio were similar or not to those in other cities integrating the same study, we reassessed in detail data from Recife, state of Pernambuco, in the Brazilian northeast,

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

located miles away from Rio, but notwithstanding one of the cities affected by the largest number of homicides in the country.[13]

Our main purpose here is to better understand one of the dimensions of RDS-associated bottlenecks, geographic bottlenecks, which, according to Khabbazian's typology corresponds to one dimension of what is called by the authors "referral bottlenecks."[6] According to our underlying assumptions, geographic bottlenecks tend to be strongly context-dependent, secondary to the influences of physical geography, access to transportation and social geography, and should be carefully assessed on a case-by-case basis. On the other hand, theoretical advances are pivotal and should incorporate simulation studies and in-depth studies of mathematical statistics. The latter is the option taken by Khabbazian et al.[6] Briefly, such complex phenomena should be the object of continuous advancements eventuating from both local, empirical studies as well as far-reaching theoretical and *in silico* experiments seeking to find new statistical and mathematical insights.

We hope our detailed assessment of Recife's processes and findings may help to better understand the role and function of geographic bottlenecks in a middle-income country where people live interacting with drug scenes.

The paper aims to analyze the tempo-ro-spatial unfolding of RDS chain-referral in context. Profiting from analyses of the successive waves over time and across different areas, we explore the potential heterogeneities stratified by Recife's geographic unities, as well as bottlenecks that might have affected the referral process and the findings eventuating from such process.

METHODS

The current study analyzes data from one of the participating cities in a multicity study ("HIV and syphilis infection rates and inventory of knowledge, attitudes, and risk practices related to sexually transmitted infections among drug users in 10 municipalities"), coordinated by the Oswaldo Cruz Foundation (Fiocruz) research team, with the support of the Department of STD, AIDS, and Viral Hepatitis, Health Surveillance Secretariat, Ministry of Health, through the International Technical Cooperation Project (AD/BRA/03/H34) between the Brazilian government and the United Nations Office on Drugs and Crime.[14,15]

The study was approved by the Institutional Review Boards of the National School of Public Health/Fiocruz headquarters (CAAE 0114.0.031.000-0) and from the regional campus located in Recife, the Aggeu Magalhães Institute/Fiocruz (CAAE 0008.0.095.031-11). All participants signed an informed consent before participation in the study, and data confidentiality was guaranteed.

The data refer to the city of Recife, Pernambuco State, Brazil, which, according to the Brazilian Institute of Geography and Statistics (IBGE), was the fourth most densely populated city of the 27 Brazilian state capitals, with a population of more than 1.5 million and a territory of 218 km^2. Recife has 94 neighborhoods, aggregated by the municipal administration in 18 micro-regions (MR) and six administrative-political regions.[16]

Six seeds (that is, potential initiators of referral chains), each residing in one of the city's administrative-political regions, were chosen during the study's preliminary phase. The seeds were four men and two women, aged 20-37 years and with different socioeconomic profiles. All six seeds reported having used crack, powder cocaine, synthetic drugs, hallucinogens, and/or tranquilizers. Only one seed (number 2) subsequently failed to recruit any other users, whereas the other five seeds recruited 394 interviewees over the course of 15 waves (recruitment rounds). According to the standard terminology of RDS studies, seed 2 was defined as a "sterile" seed, since no actual interviewee was enrolled on the basis of his contacts.

Inclusion criteria were listed as follows: legally old enough to provide informed consent (in Brazil, 18 years or older); living in the city of Recife, Pernambuco; having a valid

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

coupon (evaluated by visual inspection, then double-checked for the watermark and barcode scanned); not acutely intoxicated from recent use of alcohol and/or illegal drugs; having read (or listened to) the consent form and having signed (or fingerprinted with the right thumb, in case of illiterate individuals); and according to the CODAR criteria (for High-Risk Drug Users), having injected drugs at least once in the previous 6 months and/or having used powder cocaine, crack, opiates, methamphetamine, heroin, hallucinogens, or other illegal drugs (other than marijuana/hashish) by other routes for at least 25 days in the previous 6 months (as defined by PAHO's CODAR Manuals).[17]

Subjects that attended the assessment center received food stamps for their participation and recruitment cupons to be distributed to their peers. For each effectively recruited individual, recruiters received additional compensation (secondary incentives).

The study invitation cupons are cards that contain the public name of the survey, a short text regarding the invitation to participate, the location and hours of the center, and a bar code with the RDS number.

The interview and testing procedures took place in a downtown assessment center located in the Tamarineira neighborhood (MR 3.1), chosen due to its central location, accessibility, infrastructure, and history of user-friendly assistance provided to disenfranchised population groups.

The survey was conducted over 12 weeks (from September to December 2009) and included a socio-behavioral form delivered with an Audio Computer-Assisted Self-Interview (ACASI) system; pre- and post-test counseling aimed at reducing drug-related harms and sexual risk behaviors; and rapid testing for HIV and syphilis. Data were collected on drug use patterns in the previous 6 months and the interviewee's place of residence. As expected, there was some overlap between places of residence and places where drugs were used, since a relevant share of users had used substances at home.

For the sake of the spatial analysis of recruitment dynamics, Recife's neighborhoods (n = 94) were defined as the standard territorial unit. Digital maps, data on the resident population in each neighborhood, and the proportion of the population living in informal settings in 2010 (year of Brazil's last National Census) were obtained from the IBGE website.[18]

Informal settings correspond to the subnormal census tract defined by IBGE as

> ...any set of at least 51 housing units (e.g. shacks, huts, houses, etc.) lacking formal ownership documents and essential public services (garbage collection, electricity, sewage, public lighting) and/or arranged in a disorderly and dense way, without regular circulation/transportation routes.[19]

To depict the cumulative absolute frequency of subjects recruited over time according to the neighborhood where they lived and/or used substances, maps were drawn for each one of the five seeds that recruited at least one study participant and for the six seeds all together.

RDS recruitment trees were also drawn, depicting the process of successive recruitment of respondents over time (comprising waves or recruitment rounds).[2] Individual characteristics represented by these trees include the following information: interviewees' socioeconomic characteristics defined according to the Brazilian Economic Ranking Criteria,[20] residence in the same versus different neighborhood as the seed (dichotomous variable: "yes/no"), and residence and/or drug use in the same neighborhood as the seed (yes/no).

Poisson regression was fitted[21] to model the rate of recruited individuals per neighborhood of residence or place of drug use over time. Independent variables were divided into those primarily associated with RDS procedures themselves: distance (km) from the centroid of each neighborhood to the assessment center; presence/absence of one resident seed in a given neighborhood; presence/absence of a seed that used drugs in a given neighborhood; and those summarizing each neighborhood's demographics: proportion of the population living in informal

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

settings (multiplied by 10) and total resident population (after log transformation).

Analyses were performed using the igraph packages[22] and map tools[23] from R 3.2.2.

RESULTS

There were no refusals to answer the form among seeds (n = 6) or recruits (n = 394). However, 20% (n = 80) of recruits failed to report their place of residence and/or place of substance use in the previous 6 months. The latter subgroup included both individuals that had to live on the streets and/or obtain and use substances wherever they could find them, as well as individuals that might be suspicious about mentioning such places, despite the study's strict compliance with ethical requirements (including the protection of personal information). Unfortunately, it was not possible to distinguish between those that had no information to provide and those that did not want to provide such information.

The total number of individuals recruited through each seed (except seed 2) ranged from 21 (5.6%) through seed 6 to 214 (54.7%) through seed 3.

Seed 1 lived in the Coelhos neighborhood (MR 1.3) and reported having used drugs in the Ibura neighborhood (MR 6.2) (Figure 1). Among those recruited by the first seed (n = 70), the vast majority had been living and/or using drugs in the same neighborhood where seed 1 lived or close by, with a sizeable share that lived where the assessment center was located (especially nearby, in the Casa Amarela neighborhood - MR 3.1), as well as in the same neighborhood where seed 1 reported having used drugs, namely Ibura (Figure 1).

Seed 2 lived in the Boa Viagem neighborhood (MR 6.1) and reported having used drugs in the Cajueiro neighborhood (MR 2.2) as well as in the outskirts of the city of Recife. Since seed 2 did not recruit any participants, he was only included in the figure that describes all the seeds (Figure 1).

Seed 3 lived in the Casa Amarela neighborhood (MR 3.1) and reported having used drugs in five different neighborhoods, three of which located in MR 3.2, besides Iputinga (MR 4.1) and Pina neighborhoods (MR 6.1). Among the individuals recruited by seed 3 (n = 214), the vast majority had been living and/or using drugs in the same neighborhood where seed 3 lived, followed by those who reported having used substances in neighborhoods located close to seed 3's place of residence or close to where the assessment center was located. Seed 3's network reached the widest range of different places where people either lived or had used drugs, comprising all the administrative-political regions in Recife (Figure 1).

Seed 4 lived in the San Martin neighborhood (MR 5.1) and had been using drugs at home and in Torrões (MR 4.2) and Várzea (MR 4.3). Among the individuals recruited by seed 4 (n = 65), most lived or used drugs in the same neighborhoods where seed 4 lived/had used drugs, or very close by (Figure 1).

Seed 5 lived in the Encruzilhada neighborhood (MR 2.1) and had been using drugs at home and in Santo Amaro (MR 1.1), Arruda (MR 2.1) and Água Fria (MR 2.2). Among the individuals recruited by seed 5 (n = 24), most had been living and/or using drugs in the same neighborhood where seed 4 had lived or used drugs, or in the immediate vicinity (i.e., in the same administrative-political region).

Seed 6 lived in the Torre neighborhood (MR 4.1) and did not specify where he had been using drugs. Among the individuals recruited by him (n = 21), most had been living and/or using drugs near the assessment center, near the neighborhood where he lived (Torre), and in the city's seacoast area.

The seeds lived and used drugs in neighborhoods located in 12 of Recife's 18 micro-regions. After 15 recruitment waves, neighborhoods belonging to all the micro-regions were included either as places of residence or places of drug use. The neighborhoods with the most respondents were: Santo Amaro (MR 1.1), Coelhos (MR 1.3), Água Fria (MR 2.2), Casa Amarela (MR 3.1), and Alto José do Pinho (MR 3.2). These neighborhoods were located up to 4.1 km from the assessment center, defining a geographic cluster (considering

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

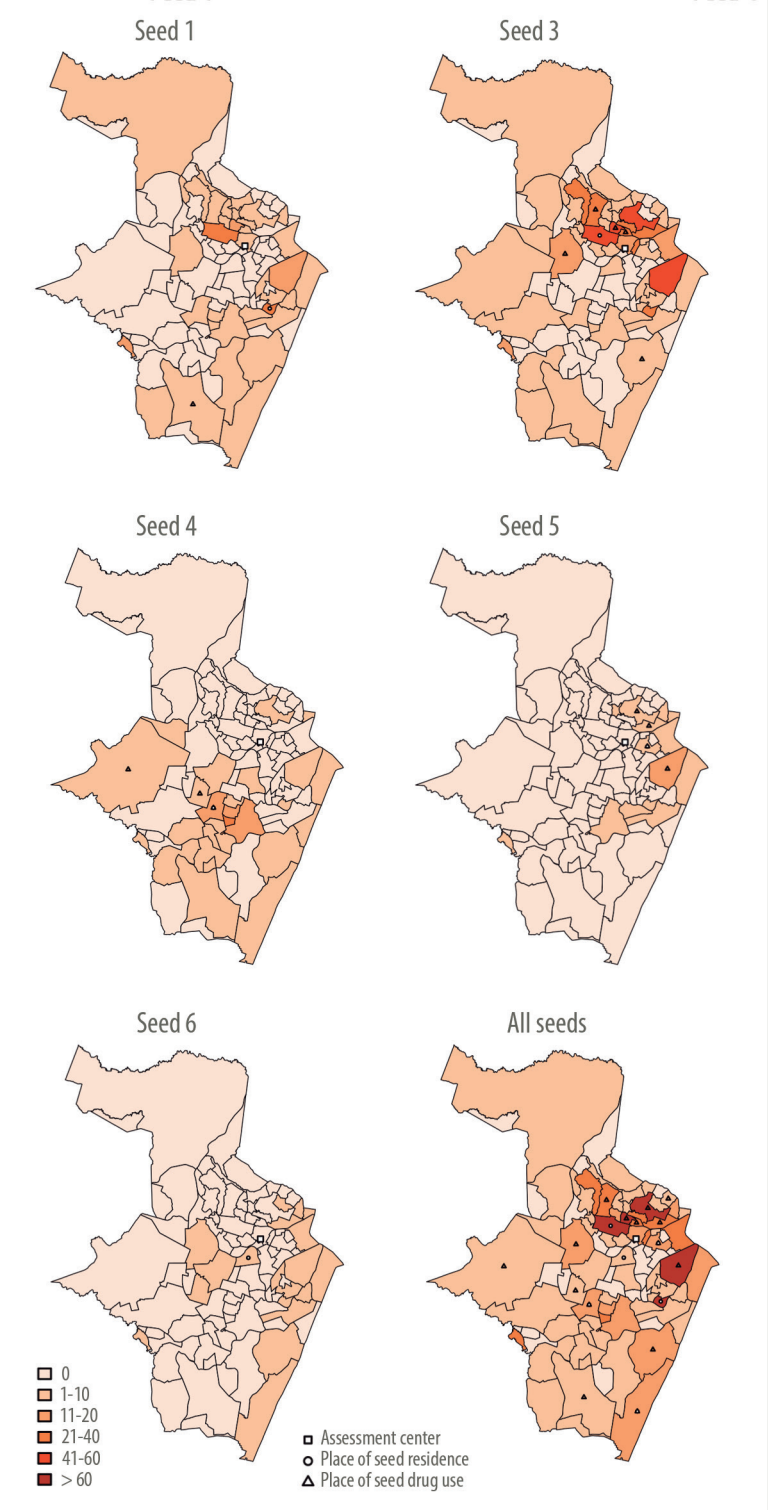

Figure 1. Cumulative frequency of subjects recruited by respondent-driven sampling (RDS), by each seed and overall, according to neighborhood of residence or of drug use. Recife, Pernambuco State, Brazil, 2009.

Source: Own elaboration.

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

a dense city with a territory of 218km^2, or 14.8 x 14.8 km if one squares its shape).

The RDS recruitment tree depicting the seeds' place of residence clearly showed this variable's strong effect (that is, pronounced homophilia, from a geographic perspective) in seed 1's network, no discernible effect on seed 6's network, and a moderate effect on the other seeds' networks (Figure 2). Notwithstanding, even in networks where the effect of seeds' place of residence proved to be "moderate," the influence of the respective seed's place of drug use and/or residence (usually as a combined variable) remained a relevant factor in the geographic concentration of recruits (Figure 3).

According to the RDS recruitment tree depicting the recruits' economic class (Figure 4), when seeds from classes "D" or "E" (the lowest ranking strata according to Brazil's standard classification) initiated the recruitment chains, it was harder for them to reach subjects belonging to higher economic strata. However, when seeds from classes "B" or "C" initiated the recruiting process, the respective referral chains tended to be more balanced, including people from diverse strata. Notwithstanding, such chains failed to reach individuals class "A." Contrary to previous findings on the influence of place of residence or of drug use, the influence of economic class was pronounced in seed 6's referral chain.

According to the adjusted model, neighborhoods D + 1 kilometers (km) from the assessment center, for any distance D, had a recruiting rate for neighborhood of residence or drug use during the study period that was 0.81 [95% CI (0.79-0.84)] times that of the recruiting rate for neighborhoods located D km from the assessment center. That is, the

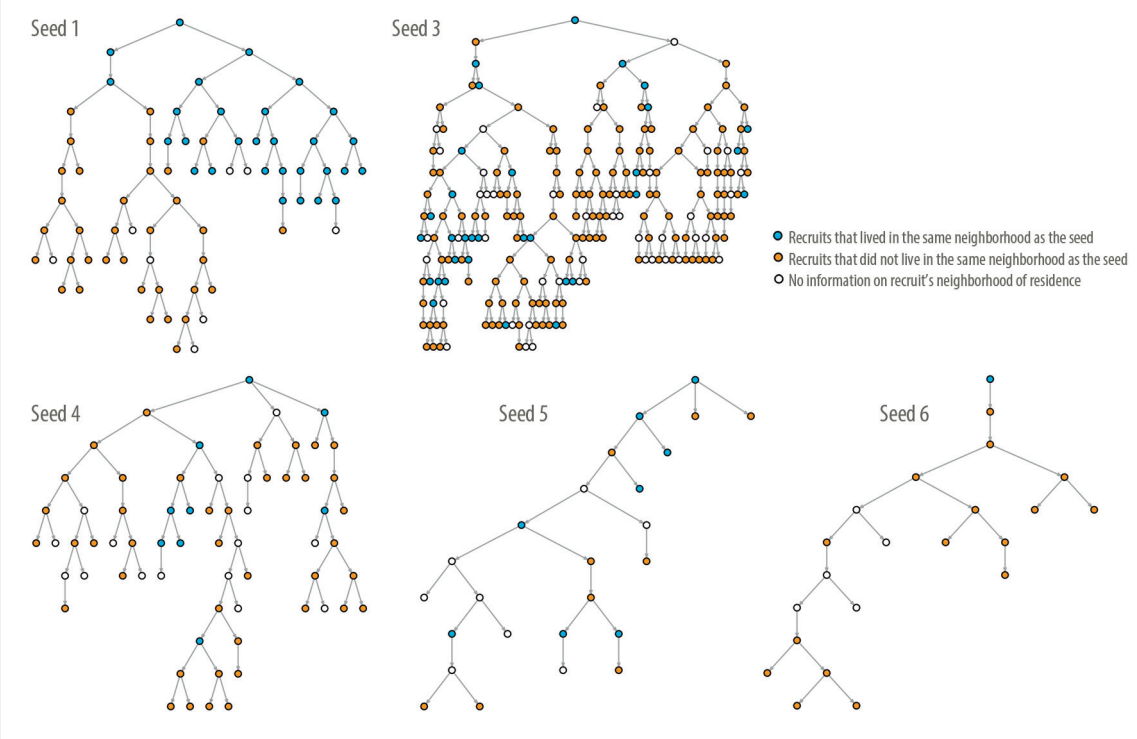

Figure 2. Respondent-driven sampling (RDS) recruitment tree according to the place of residence of recruit versus respective seed. Recife, Pernambuco State, Brazil, 2009.

Source: Own elaboration.

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

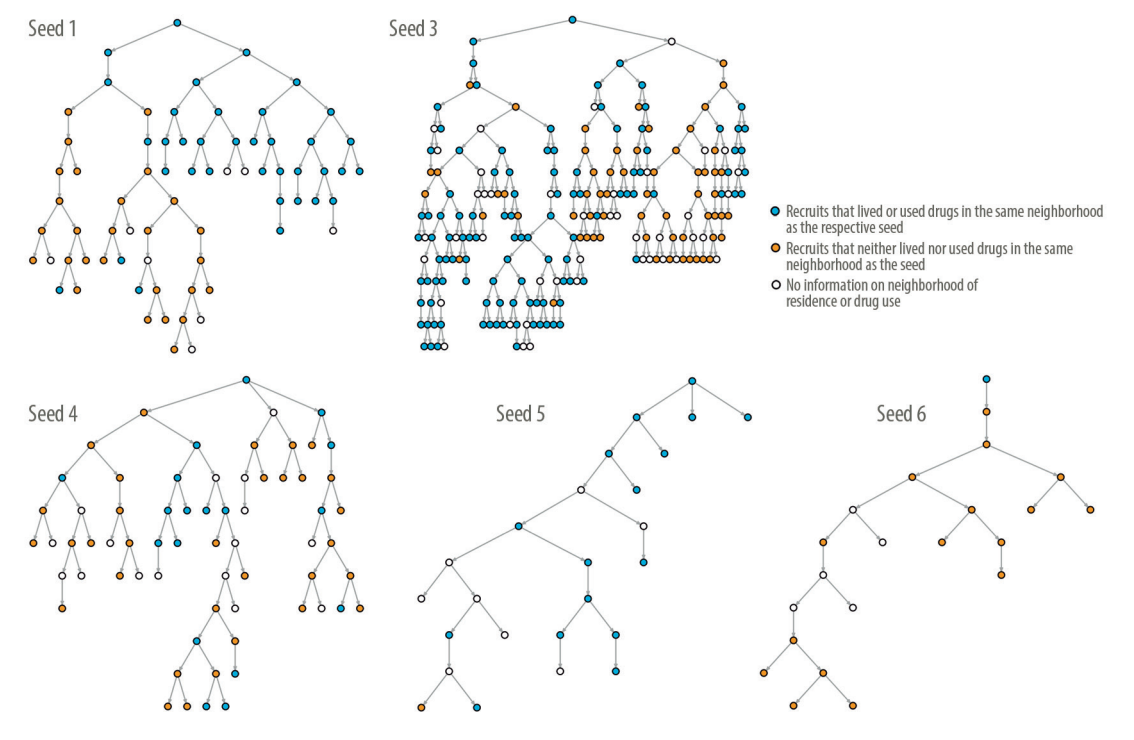

Figure 3. Respondent-driven sampling (RDS) network recruitment tree according to place of residence or drug use in the same neighborhood as the seed's drug use or residence. Recife, Pernambuco, State, Brazil, 2009.

Source: Own elaboration.

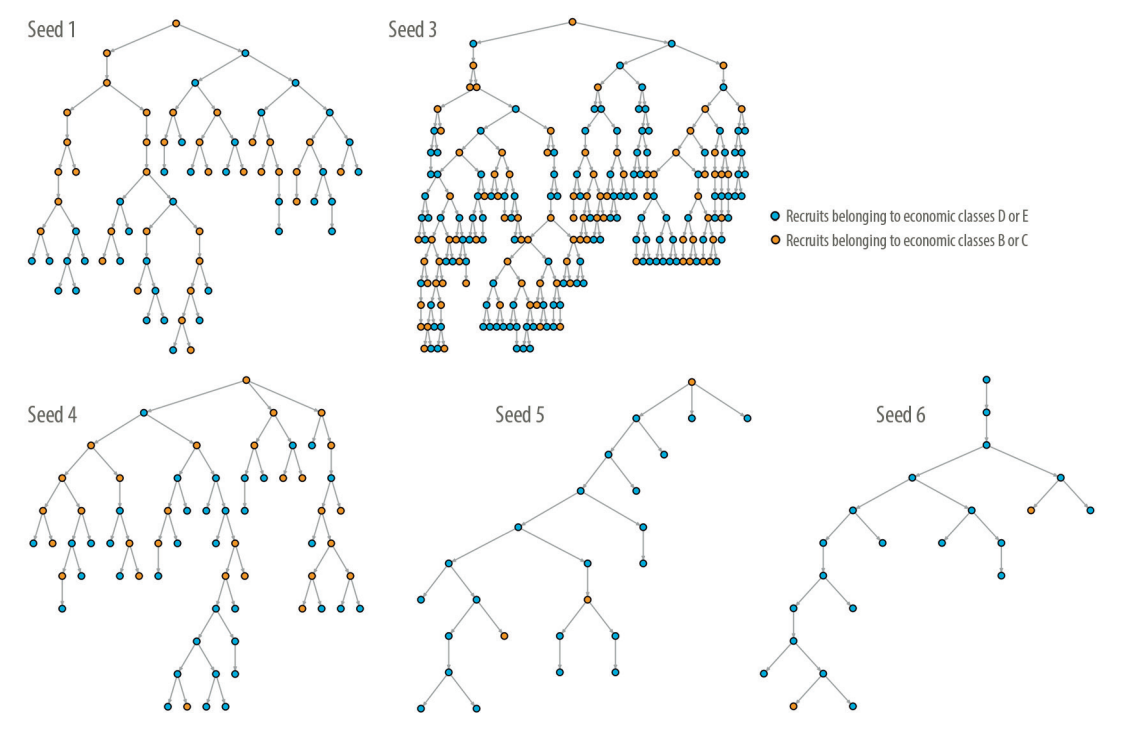

Figure 4. Respondent-driven sampling (RDS) network recruitment tree according to participants' economic class. Recife, Pernambuco State, Brazil, 2009.

Source: Own elaboration.

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

Table 1. Adjusted analysis of demographic and spatial factors associated with the number of recruits per neighborhood of residence (n=94 neighborhoods) or drug use during the study period. Recife, Pernambuco State, Brazil, 2009.

Demographic and spatial factors	Ratio	95% CI
Distance from neighborhood to the assessment center (km)	0.81	0.79-0.84
Seed lived in the neighborhood		
Yes	2.46	2.08-2.89
No[1]	1	-
Seed used drugs in the neighborhood		
Yes	2.70	2.33-3.13
No[1]	1	-
Proportion of population in informal settings (x 10)	1.10	1.08-1.13

Source: Own elaboration.
Note: Controlled by total population of the neighborhoods according to 2010 Census data.
[1]Reference value.

farther the neighborhood from the assessment center, the lower the rate of recruits per neighborhood (Table 1).

The recruiting rate per neighborhood of residence or drug use during the study period in the neighborhoods where the seeds lived was 2.46 times [95% CI (2.08-2.89)] higher than other neighborhoods, and was 2.70 times [95% CI (2.33-3.13)] higher in neighborhoods where the seeds used drugs than in neighborhoods where they did not use drugs (Table 1).

For each increase in population living in informal settings of 10%, the recruiting rate per neighborhood of residence or drug use during the study period increases 1.10 times [95% CI (1.08 -1.13)] (Table 1).

DISCUSSION

Bottlenecks may be associated with several varieties of social heterogeneity, inequality and prejudice.

Ideally, different social strata defined by age, gender, ethnicity, income, social class should be connected by some kind of bridging, that is, groups of individuals who may function as "brokers" between such segments.

Unfortunately, due to a series of factors, such as structural violence,[8] generation and income gaps,[24] poorly connected segments such as recently arrived immigrants,[25] as well as entrenched prejudice against some populations such as transgender women,[26] this does not happen in several real-life studies (as commented by Salganik[7]).

Heterogeneities and consequent bottlenecks frequently have an explicit spatial dimension, being the context rural Uganda[10] or a major urban area, such as Rio de Janeiro.[8] Since spatial heterogeneities and respective recruiting bottlenecks are not necessarily clear at a first sight, we carried out a detailed analysis of the real-life implementation of RDS in Recife, Pernambuco, Brazil.

The analysis of the recruitment distribution using maps, according to the neighborhoods where individuals lived and/or used drugs, showed that in general the seeds' networks tended to include more individuals living or using drugs in the same neighborhoods as the seeds, nearby, or located near the assessment center. Briefly, whatever the meaning people ascribe to the concepts of proximity and/or accessibility, they appear to be key determinants of the recruiting dynamic.

The recruitment trees showed that the effect of the neighborhood where the seed lived and/or used substances (usually in combination) was pivotal in the recruitment process, even in the context of networks in which the effect of vicinity was not as pronounced as in other networks. Despite the high concentration of participants in economic classes D and E, the sample also included individuals from classes B and C. However, class A (the elite) remained out of reach, as an entirely invisible stratum.

The Poisson regression model corroborated these findings. The model showed that increasing the distance from each neighborhood to the assessment center resulted in a lower recruiting rate in this specific place. Neighborhoods where the seeds had been living and/or using drugs had higher recruiting rates during the study period, and there was a positive association between the proportion of the population living in informal clusters in neighborhoods and the recruiting rate per

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

neighborhood where individuals had been living and/or using drugs during the study period.

Studies conducted in other settings have also highlighted the importance of the assessment center's location and the place where individuals perform their daily activities.

According to a study in 2009 with a sample of 605 drug users living in the city of Rio de Janeiro, Brazil, the interviewees clustered around the assessment center, where most of the seeds lived.[8]

Another study in 2006-2007 with 1,048 injecting drug users in Tijuana, Mexico, showed that individuals who lived, worked, injected and/or bought drugs in the North Zone (known for its sex and drug trades) tended to recruit others individuals who had engaged in activities in the same neighborhood. This finding documents the important effect of places where individuals live and interact on the spatial clustering of recruitment.[27]

A study in 2009-2010 with a cohort of 2,402 men living in 25 villages in rural Uganda showed that although all the towns were represented in the sample, individuals living more than 1 km from the assessment center showed nearly 50% lower odds of being recruited into the RDS study. Even individuals who were known to have received cupons but who lived more than 1 km from the assessment center showed approximately 30% lower odds of appearing at the assessment center.[10]

The recruitment bias produced by the selection of geographically proximate pairs can lead to over- or under-recruitment of individuals that live and/or interact socially in the same or nearby neighborhood. In this case, a group can be overrepresented or underrepresented in the RDS sample, and the network may include individuals who have characteristics that are more similar than observed in the reference population. These recruitment patterns can introduce bias in the estimates and erroneously narrow the confidence intervals, especially in geographically heterogeneous populations.[27]

As expected, a large proportion of participants were recruited from economic classes D and E when compared to classes B and C and especially the "silent" elite A class. Given the deep social and demographic heterogeneities between neighborhoods with higher proportions of the population living in informal clusters and other neighborhoods, such specificities would likely interact and perhaps be superimposed on the drug user population's general characteristics. Similar conclusions have been reached in various countries where there is marked social heterogeneity and severe marginalization of certain populations (such as illegal drug users), which is especially common in Latin America[28] and throughout Eastern Europe.[29]

This complex geographic dynamic is present even in high-income countries such as the United States, albeit with specific characteristics of ethnic and linguistic contrasts, as well as the establishment of geographic enclaves or segregated areas.[30]

The study of men living in villages in rural Uganda showed that participants tended to recruit lower-income individuals rather than selecting pairs randomly chosen from their contact network,[24] showing that recruitment/selection was far from random within each social network. We cannot rule out the hypothesis that drug users recruited in the current study selected their lower socioeconomic-class peers within their respective networks, especially due to the incentives to recruit other individuals and the compensation offered for participation in the study.

An interesting observation from the study was that seeds from higher economic classes tended to generate more heterogeneous recruitment networks. The reverse direction has evident barriers, involving the difficulty of individuals from disenfranchised classes in recruiting networks with greater social diversity. This fact has an obvious numerical dimension, since the affluent segments of Brazilian society are tiny compared to the numerically larger low and middle classes. The poorest segments of Brazilian society face deeply entrenched prejudice, which hinders their interaction with segments displaying higher income, better education, and well-paid occupations. These affluent segments tend to be white and have cosmopolitan

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

habits and mores, as documented by Brazilian and international sociology.[31]

A limitation of this study, similar to that of other articles addressing the same issues,[10], was that distances were calculated along straight lines (simple Euclidean distances using the neighborhoods' centroids) due to the lack of addresses and knowledge of the actual paths traveled by individuals as they commute between different locations.

Another important limitation was the lack of information on the neighborhoods where substantial proportions of the drug users actually lived and consumed drugs, as observed by Toledo et al.[8] The limitation appears to derive from this population's characteristics and the key role of structural violence in Brazilian society, especially in recent years.[9] That is, it may or may not be *possible* to locate individuals, or they may or may not *want* to be located, given the permanent threat of stigmatization and criminalization of their habits and behaviors.

Brazilian legislation lacks a clear distinction between drug users and dealers, as well as objective criteria for defining personal use, coupled with severe penalties for trafficking. Theses factor contribute to the growing incarceration of individuals from poor areas (especially poorly educated minority youth) and to the obvious reluctance to provide personal data, which can dramatically increase their odds of arrest and/or imprisonment.[32]

Another limitation deserved mention. The existence of a single assessment center may have led to a clustering of interviewees around the center, located near administrative-political region 3. It may have also influenced the number of participants recruited by each seed, with the largest share of the sample derived from a single recruitment chain (originating from seed 3, who lived in administrative-political region 3). McCreesh et al.[10] also emphasize the importance of careful choice of the assessment center for the success of RDS sampling.

Despite the above-mentioned limitations and the observed spatial bias, the sample obtained via chains of successive RDS waves successfully reached individuals that lived or used drugs in all of Recife's micro-regions and four of the five socioeconomic classes (B to E, but not "A"), showing that the RDS method is a valid alternative for recruiting illegal drug users in Recife. In contemporary Brazil, the delivery of customized batches of illegal substances via selected vendors, mobile phone networks and the internet has been growing, and thus socioeconomic class-A users may interact little or not at all with traffic scenes. In this sense, bridges between these and other drug users would be rare or nonexistent.[33]

However, any discussion of the validity of the estimates obtained from this sample should be made in light of the outcomes to be estimated and their association with the variables for which homophilia may be a relevant issue. The higher the homophilia, the lower the odds of generating new information with each wave. In a hypothetical case of perfect homophilia, the referral chain would indefinitely spin around the same characteristics of the seeds, following a clonal process, precluding any valid inference concerning the target population.[2]

Although RDS was originally defined as a process "without memory," based on a first-order Markov chain,[1,2] this was not seen in the current study, where the geographic dimension was key.

The study highlights factors that may favor successful assessment (via RDS) of the geographic spread of any infectious disease across drug user chains, as follows: 1) selection of seeds living and/or using drugs in as many neighborhoods as possible, seeking to contemplate a given city's various micro-regions and neighborhoods; 2) selection of more than one assessment center in different regions; and 3) inclusion of seeds from higher socioeconomic classes (although inclusion of the more exclusive segments of the elite is a practically unachievable goal).

CONCLUSIONS

Recruitment chains depend on the social geography and demographics of the population. A

SALUD COLECTIVA. 2020;16:e2524. doi: 10.18294/sc.2020.2524

well-balanced study should incorporate seeds residing and/or using drugs in as many neighborhoods as possible. Ideally, the coordinated implementation of more than one assessment center (located in different regions) and the intentional inclusion of seeds from higher socioeconomic strata (in order to yield more heterogeneous networks) should be attempted, aiming to improve RDS studies of this population in real-world conditions.

One of the main issues to be addressed by RDS-based studies are bottlenecks that preclude the enrollment of potential recruits to match the goal of comprehensiveness and, ideally, the exhaustive mapping of a given hard-to-reach population. Bottlenecks determined by geography are just one among several others, such as racism, sexism, lack of trust between different generations of a given population/community. Notwithstanding, since most people live in large urban areas, in Brazil and several rich and low/middle-income countries worldwide, the proper assessment of geographic bottlenecks and ways to handle them in the best way possible is key.

ACKNOWLEDGEMENTS

The study analyzes data from one of the participating cities in a multicity study "HIV and syphilis infection rates and inventory of knowledge, attitudes, and risk practices related to sexually transmitted infections among drug users in 10 municipalities," coordinated by the Oswaldo Cruz Foundation (Fiocruz) research team, with the support of the Department of STD, AIDS, and Viral Hepatitis, Health Surveillance Secretariat, Ministry of Health, through the International Technical Cooperation Project (AD/BRA/03/H34) between the Brazilian government and the United Nations Office on Drugs and Crime. Sandro Sperandei is the recipient of a postdoctoral scholarship from Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

1. Heckathorn DD. Respondent driven sampling: a new approach to the study of hidden population. Social Problems. 1997;44(2):174-199. doi: 10.2307/3096941.

2. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. Social Problems. 2002;49(1):11-34. doi: 10.1525/sp.2002.49.1.11.

3. Gile KJ, Beaudry IS, Handcock MS, Ott MQ. Methods for inference from respondent-driven sampling data. Annual Review of Statistics and Its Application. 2018;5(1):65-93. doi: 10.1146/annurev-statistics-031017-100704.

4. Johnston LG, Hakim AJ, Dittrich S, Burnett J, Kim E, White RG. A systematic review of published respondent-driven sampling surveys collecting behavioral and biologic data. AIDS and Behavior. 2016;20(8):1754-1776. doi: 10.1007/s10461-016-1346-5.

5. Montealegre JR, Johnston LG, Murrill C, Monterroso E. Respondent driven sampling for HIV biological and behavioral surveillance in Latin America and the Caribbean. AIDS and Behavior. 2013;17(7):2313-2340. doi: 10.1007/s10461-013-0466-4.

6. Khabbazian M, Hanlon B, Russek Z, Rohe K. Novel Sampling Design for Respondent-driven Sampling. Electronic Journal of Statistics. 2017;11(2):4769-4812. doi: 10.1214/17-EJS1358.

7. Salganik MJ. Commentary: Respondent-driven Sampling. Epidemiology. 2012;23(1):148-150. doi: 10.1097/EDE.0b013e31823b6979.

8. Toledo L, Codeço CT, Bertoni N, Albuquerque E, Malta M, Bastos FI, et al. Putting respondent-driven sampling on the map: insights from Rio de Janeiro, Brazil. Journal of Acquired Immune Deficiency Syndromes. 2011;57(Suppl 3):S136-S143. doi: 10.1097/QAI.0b013e31821e9981.

9. Bastos FI. Structural violence in the context of drug policy and initiatives aiming to reduce drug-related harm in contemporary Brazil: a review. Substance Use & Misuse. 2012;47(13-14):1603-1610. doi: 10.3109/10826084.2012.705678.

10. McCreesh N, Johnston LG, Copas A, Sonnenberg P, Seeley J, Hayes RJ, et al. Evaluation of the role of location and distance in recruitment in respondent-driven sampling. International Journal of Health Geographics. 2011;10(1):56. doi: 10.1186/1476-072X-10-56.

11. Young AM, Rudolph AE, Quillen D, Havens JR. Spatial, temporal and relational patterns in respondent-driven sampling: evidence from a social network study of rural drug users. Journal of Epidemiology and Community Health. 2014;68(8):792-798. doi: 10.1136/jech-2014-203935.

12. Montealegre JR, Risser JM, Selwyn BJ, McCurdy SA, Sabin K. Effectiveness of Respondent Driven Sampling to Recruit Undocumented Central American Immigrant Women in Houston, Texas for an HIV Behavioral Survey. AIDS and Behavior. 2013;17(2):719-727. doi: 10.1007/s10461-012-0306-y.

13. Waiselfisz JJ. Mapa da Violência 2016: Homicídios por Armas de Fogo no Brasil. Brasilia: FLACSO Brasil; 2016.

14. Bastos F. Taxas de infecção de HIV e sífilis e inventário de conhecimento, atitudes e práticas de risco rela-

cionadas às infecções sexualmente transmissíveis entre usuários de drogas em 10 municípios brasileiros. Brasília: Ministério da Saúde, Departamento de Vigilância, Prevenção e Controle das IST, do HIV/Aids e das Hepatites Virais; 2009.

15. Santos NTV. Vulnerabilidade e prevalência de HIV e sífilis em usuários de drogas no Recife: resultados de um estudo respondent-driven sampling. [Tese de Douturado]. Recife: Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz; 2013.

16. Instituto Brasileiro de Geografia e Estatística. Cidades [Internet]. 2016 [cited 23 Mar 2018]. Available from: https://tinyurl.com/y59us2kz.

17. Organización Panamericana de la Salud. Encuesta de comportamiento en CODAR: Herramientas básicas Diseño del estudio adaptación del cuestionario e indicadores. Washington DC: OPS; 2008.

18. Instituto Brasileiro de Geografia e Estatística. Downloads [Internet]. 2016 [cited 23 Mar 2018]. Available from: https://tinyurl.com/ya74669m.

19. Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2010: Aglomerados Subnormais - Primeiros Resultados. Rio de Janeiro: IBGE; 2011.

20. Associação Brasileira de Empresas de Pesquisa. Critério Brasil [Internet]. [cited 23 Mar 2018]. Available from: http://www.abep.org/criterio-brasil

21. McCullagh P, Nelder JA. Generalized Linear Models. 2nd ed. New York: Chapman & Hall, CRC; 1989.

22. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Systems. 2006;1695.

23. Bivand R, Lewin-Koh N, Pebesma E, Archer E, Baddeley A, Bibiko H-J, et al. Maptools: Tools for handling spatial objects [Internet]. 2014 [cited 23 mar 2018]. Available from: https://tinyurl.com/y36pw4xa.

24. McCreesh N, Frost SDW, Seeley J, Katongole J, Tarsh MN, Ndunguse R, et al. Evaluation of respondent-driven sampling. Epidemiology. 2012;23(1):138-147. doi: 10.1097/EDE.0b013e31823ac17c.

25. Frere-Smith T, Luthra R, Platt L. Sampling recently arrived immigrants in the UK: Exploring the effectiveness of respondent driven sampling. Colchester: University of Essex, Institute for Social and Economic Research; 2014. ISER Working Paper Series, No. 2014-25.

26. Bastos FI, Bastos LS, Coutinho C, Toledo L, Mota JC, Velasco-de-Castro CA, et al. HIV, HCV, HBV and syphilis among transgender women from Brazil: Assessing different methods to adjust infection rates of a hard-to-reach, sparse population. Medicine (Baltimore). 2018;97(Suppl 1):S16-S24. doi: 10.1097/MD.0000000000009447.

27. Rudolph AE, Gaines TL, Lozada R, Vera A, Brouwer KC. Evaluating outcome-correlated recruitment and geographic recruitment bias in a respondent-driven sample of people who inject drugs in Tijuana, Mexico. AIDS and Behavior. 2014;18(12):2325-2337. doi: 10.1007/s10461-014-0838-4.

28. Rossi D, Zunino Singh D, Pawlowicz MP, Touzé G, Bolyard M, Mateu-Gelabert P, et al. Changes in time-use and drug use by young adults in poor neighbourhoods of Greater Buenos Aires, Argentina, after the political transitions of 2001-2002: Results of a survey. Harm Reduction Journal. 2011;8(1):2. doi: 10.1186/1477-7517-8-2.

29. Rhodes T, Ball A, Stimson G V, Kobyshcha Y, Fitch C, Pokrovsky V, et al. HIV infection associated with drug injecting in the newly independent states, eastern Europe: the social and economic context of epidemics. Addiction. 1999;94(9):1323-1336.

30. Galea S, Rudenstine S. Challenges in understanding disparities in drug use and its consequences. Journal of Urban Health. 2005;82(Suppl 3):S5-S12. doi: 10.1093/jurban/jti059.

31. Velez C, Barros R, Ferreira F. Inequality and economic development in Brazil. Washington DC: World Bank Publications; 2014.

32. Boiteux L. Drugs and prisons: The repression of drugs and the increase of the Brazilian penitentiary population. In: Metaal P, Youngers C, (eds.). Systems overload: Drug laws and prisons in Latin America. Amsterdam: Transnational Institute, the Washington Office on Latin America; 2011.

33. Soares L. Tudo ou nada. Rio de Janeiro: Nova Fronteira; 2012.

CITATION
Valois-Santos NT, Niquini RP, Sperandei S, Bastos LS, Bertoni N, Brito AM, Bastos FI. Reassessing geographic bottlenecks in a respondent-driven sampling based multicity study in Brazil. Salud Colectiva. 2020;16:e2524. doi: 10.18294/sc.2020.2524.

Received: 31 Aug 2019| Modified: 17 Jun 2020 | Accepted: 15 Jul 2020 | Publication online:

https://doi.org/10.18294/sc.2020.2524.