Spatial distribution of HIV prevalence and incidence among injection drugs users in St Petersburg: implications for HIV transmission

Robert Heimer\textsuperscript{a}, Russell Barbour\textsuperscript{a}, Alla V. Shaboltas\textsuperscript{b}, Irving F. Hoffman\textsuperscript{c} and Andrei P. Kozlov\textsuperscript{b}

\textbf{Introduction:} The HIV/AIDS epidemic in St Petersburg, as in much of Russia, is concentrated among injection drug users (IDU) in whom prevalence reached 30\% in 2003. Understanding the dynamics of the epidemic is important in developing appropriate responses in the resource-constrained context of Russian cities such as St Petersburg.

\textbf{Methods:} IDU were contacted and screened to create a seronegative cohort for prevention and vaccine studies. At screening, individuals provided sociodemographic, drug use, and injection and sex-related risk behavior data. Seronegative individuals who enrolled in the cohort were followed for one year and tested for HIV semiannually. Residential addresses were entered into a geographical information system programme and analysed for spatial clustering using Moran’s $I$ and nearest-neighbor analysis.

\textbf{Results:} We mapped 788 of the 900 study participants to discrete locations within St Petersburg; 236 (29.9\%) were HIV seropositive at baseline. Although there was no clustering of the study population as a whole, HIV-infected individuals were tightly clusterd and prevalence co-clustered with high frequency of heroin injection, receptive syringe sharing, being younger than 24 years, and living with parents. These clusters were restricted to 5\% of populated areas of the city. We mapped 18 of 20 incident cases detected among the cohort, and more than half were located within or adjacent to the clusters.

\textbf{Interpretation:} Spatial analysis identified linkages between disease prevalence and risky injection behaviors that were not evident using traditional epidemiological analysis. The analysis also identified where resources might be allocated geographically for maximum impact in slowing the HIV epidemic among IDU.

\textbf{Keywords:} geographical information systems, harm reduction, HIV/AIDS, injection drug use

Introduction
The Russian Federation has witnessed one of the fastest growing HIV epidemics in the world over the past decade [1–3]. The number of registered HIV infections has increased from 1090 at the end of 1995 to more than 330,000 by the end 2005. The city of St Petersburg has been particularly hard hit. By the end of 2005, nearly...
29,000 infections had been recorded among its residents, the most for any of the 90 administrative regions within Russia. As is true elsewhere in Russia, most of the HIV infections have occurred in individuals who reported engaging in injection drug abuse. From 1996 to 2001, as many as 90% of infections were attributed to injection drug use [2,4]. During the 2 years covered by this study, 2002–2003, 72 and 54% of newly reported HIV infections were in individuals reporting injection drug use; similar percentages were found for 2004 and 2005, 65 and 66%, respectively [5]. To understand and control the HIV epidemic in places such as St Petersburg it is vital to understand how it is continuing to spread among injectors.

In 2002, we began a study to collect data from active drug injectors in order to recruit a sample of at least 500 individuals into a cohort to determine HIV seroincidence. A total of 900 individuals were screened for HIV based on meeting the study criteria for active injection drug use and 270 (30.0%) were found to be HIV positive [6]. Analysis of the baseline data to identify factors associated with seroprevalence found no significant associations of behavioral risk factors with prevalence, and the only significant association of prevalence with demographic factors was being younger, but the difference in mean and median age between the two groups was only approximately one year. Of those who were seronegative, 520 enrolled in the cohort study, agreeing to make two semi-annual visits for repeat testing. Despite the provision of HIV risk-reduction counseling and free medical care, 20 infections were detected among the 80% who were retained in the cohort, an annualized incidence rate of 4.5 per 100 person-years of exposure [7]. Analysis of data collected from cohort members who returned for follow-up found only one significant risk factor associated with seroconversion, the injection of ephedrine-derivative psychostimulants. We were unable to find a common thread that united existing and emergent infections in our study population.

To generate a more sophisticated analysis of the spread of HIV among drug injectors in St Petersburg, we undertook a spatial analysis of the prevalent and incident HIV infection. We were able to geolocate 788 of the 900 individuals screened using the home addresses provided at screening. This analysis represents a unique example of the power of spatial analysis to integrate otherwise unrelated factors to create a unified picture of an HIV epidemic in a city and to provide the necessary documentation for a rational public health response that will confront the HIV epidemic where it is most expansive and relevant.

Methods

Cohort recruitment and HIV testing
As part of the HIV Prevention Trials Network (HPTN) Protocol 033 that assessed seroincidence in four sites with burgeoning HIV epidemics driven in significant measure by injection drug use, 900 active injection drug users (IDU) in St Petersburg, Russia, were recruited, interviewed, and tested for HIV between January and December 2002. Individuals were tested using the Vironostika HIV-UniForm II plus O enzyme-linked immunosorbent assay (BioMerieux, the Netherlands) and positive test results were confirmed by Western blot (New Lav Blot 1; Bio–Rad Laboratories, France). Of these, 520 who tested seronegative were entered into a cohort to be followed for one year to detect new HIV infections. Individuals were provided risk reduction counseling appropriate to their test results, offered free primary healthcare, and compensated with gifts valued at approximately US$5 for each study visit. Follow-up visits were scheduled semi-annually. The protocols for recruitment, testing and counseling, and follow-up were approved by institutional review boards at the University of North Carolina and the Biomedical Center, St Petersburg.

Collection of demographic and behavioral data
Data were collected at baseline using an interview questionnaire common to all four HPTN033 sites. Drug use questions focused on drugs injected lifetime and the past 30 days, injection frequency, and injection risk behaviors, including syringe and non-syringe paraphernalia sharing. Sexual risk was investigated through questions concerning lifetime and past 6 month behaviors. Locator information was collected from each of the 900 participants in the form of their residential address.

Spatial analysis
Spatial analysis was initiated by geolocating residential addresses of respondents using St Petersburg city electronic maps (Institute of Applied Geography, St Petersburg, Russian Federation). Of 900 addresses in the original survey, 112 locations were outside of the area of study, could not be located, or could not be unambiguously mapped. The remaining 788 locations within St Petersburg were mapped using ArcGIS 9.1 (ESRI Corp., Redlands, California, USA). We confirmed that all addresses were located within residential areas and not at non-residential locations such as rail yards or industrial areas.

We then explored the data for significant spatial clustering of demographic and HIV risk behaviors. Spatial analysis, as applied in this study, is derived from earlier efforts to formalize exploratory data analysis [8]. The availability of geographic information systems software has facilitated this process. Spatial autocorrelation, i.e. the correlation of a single variable between pairs of neighboring observations, is the central concept of this form of exploratory data analysis [9]. The term ‘neighboring observation’ is user-defined by creating a contiguity or distance matrix in which the correlation between neighbors is compared with the general variance of the sample as in ordinary correlation analysis. Moran introduced what is now the standard measure of spatial autocorrelation, ‘Moran’s
It estimates the strength of correlation between observations as a function of the distance separating them. CRIMESTAT III software employed in this study calculates Moran’s $I$ using the following equation:

$$ I = \frac{\sum_{i,j} W_{ij} (z_i - \bar{z}) \cdot (z_j - \bar{z}) / \sigma^2(z)}{n} $$

where $\sigma^2$ is the sample variance and $W_{ij}$ is the proximity matrix [11]. Proximity matrix values are 1 only for those locations $i$ and $j$ that are contiguous, all other values are zero. Similar to Pearson’s correlation coefficient, the values of Moran’s $I$ range from +1 (indicating strong autocorrelation) through 0 (indicating a random pattern) to −1 (indicating over-dispersion and uniformity). Oliveau [9] noted that strong negative values are extremely unusual in demographic and social science data. The CRIMESTAT III algorithm assigns a $P$ value to the calculated Moran’s $I$ value permitting inference as to the degree and significance of spatial clustering compared with an expected random value.

Clusters were identified using nearest-neighbor analysis, a single linkage, hierarchical method in CRIMESTAT III [11]. It uses a constant-distance clustering routine that groups points together on the basis of spatial proximity [12]. The user defines a threshold distance and the minimum number of points that are required for each cluster, and an output size for displaying the clusters with ellipses. The routine identifies first-order clusters, representing groups of points that are closer together than the threshold distance and in which there is at least the minimum number of points specified by the user. Clustering is hierarchical in that the first-order clusters are treated as separate points to be clustered into second-order clusters, and second-order clusters are treated as separate points to be clustered into third-order clusters, and so on. Higher-order clusters are identified only if the distance between their centers is closer than the new threshold distance [11]. In this application only first order clusters were used because no significant higher order clusters were found. In this application we set the minimum number of points to be included in a cluster at 40, the threshold distance to one kilometer for all variables, and the alpha value to 0.05.

Results were graphically displayed as ellipses by exporting ArcGIS shape files generated in CRIMESTAT III to ArcGIS. Ellipse size was set at one standard deviational ellipse of the cluster (the rotation and the lengths of the $X$ and $Y$ axes) because this specification allowed us to exclude non-residential areas of the city such as rail yards and waterways, yet maintain the statistical significance of clusters contained within known residential areas.

The residential areas within St Petersburg are not randomly arranged, and therefore we used a programme called HawthTools to generate random points in continuous space within the polygons of residential areas [13]. The tool generates points with a uniform distribution for $x$ and $y$ coordinates in the larger spatial domain of interest, which in our case was the residential districts of the city of St Petersburg. HawthTools then determines whether to keep each point based on whether or not it falls within the polygons of interest. The randomly generated points were analysed to exclude the possibility that random points in the residential areas would be clustered because of the configuration of these areas.

### Results

We determined that even though the residential areas of St Petersburg have a discrete spatial organization, it was possible to distribute points randomly equivalent to the number of study participants throughout these areas without significant clustering. We then mapped the spatial distribution of the 788 residential addresses within St Petersburg. These were distributed throughout St Petersburg in a pattern more closely resembling random than clustered (Moran’s $I = 0.049$, $P > 0.1$).

Individuals were recruited using different methods at different stages of the one-year recruitment period [6], and therefore we explored the influence of recruitment order on the spatial distribution of individuals recruited into the study. The first 160 geocoded individuals were recruited primarily by venue-based sampling, identifying active IDU at drug treatment programmes. The next 320 geocoded individuals were recruited through a combination of location-based screening, street outreach, and chain referrals from existing participants. Thereafter, chain referral was the predominant method of recruitment. Significant clustering was observed after the recruitment of 160, 320, 480, and 640 individuals, a finding in keeping with the reliance on location-based sampling methods at the outset of the study (Table 1). A pattern more closely resembling random emerged after all individuals were recruited (Fig. 1).

In contrast, the locations of residence for individuals with prevalent HIV infection were found to be strongly

<table>
<thead>
<tr>
<th>Number mapped</th>
<th>Moran’s $I$</th>
<th>$P$ value of spatial clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>First 160</td>
<td>0.078 ± 0.012</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>First 320</td>
<td>0.153 ± 0.012</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>First 480</td>
<td>0.132 ± 0.012</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>First 640</td>
<td>0.057 ± 0.012</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>All 788</td>
<td>0.048 ± 0.012</td>
<td>NS</td>
</tr>
</tbody>
</table>

*The Moran’s $I$ values were calculated by comparison with 900 points generated randomly within the residential areas of St Petersburg.*
clustered (Fig. 2a, Table 2). Four spatial clusters within the city were found to contain 72 of the 236 mapped cases. Clustering was also detected for several of the injection practices. Almost all injectors reported injecting heroin; only 35 of 900 (3.9%) reported not doing so in the month before their baseline interview. Individuals injecting heroin only three or more times per week (Table 2) were more likely to reside in one of five clusters (Fig. 2b, Table 2). Psychostimulant injection was reported by 299 individuals (37.9%). These injectors were not significantly associated with HIV prevalence or incident cases. Heroin frequency and needle sharing were significantly clustered (Fig. 2c, Table 2). None of the risky sexual behaviors addressed in the risk assessment survey was correlated with HIV prevalence. Living with parents and being younger than 24 years were correlated, but living with parents is not correlated with HIV prevalence and neither was correlated with injection frequency or syringe sharing.

The significant clusters were confined to small regions within the residential regions of the city, which covers 37.13 square kilometers of the 144.27 square kilometers of land within the city limits. The total cluster size for the four significantly clustering variables, HIV prevalence, living with parents, high heroin injection frequency, and receptive syringe sharing, ranged from 0.3 to 4.6% of the total populated area (Table 2). As is evident in Fig. 3, many of these clusters overlapped. The most significant overlaps were located near the Sitniy market on Petrograd Island and surrounding the Ladozhskaya metro stop on the east side of the Neva River.

We were able to map residences for 18 of the 20 individuals who became infected with HIV while enrolled in the study. These incident cases were themselves not significantly clustered, a finding not surprising because there are so few observations. Nevertheless, 12 of the 18 were located in or near one or more of the clusters. Ten were located in the overlap of at least three clusters, six in the Sitniy cluster and four in the East Neva cluster (Fig. 4).

The analysis presented in this report is a unique combination of spatial, demographic, behavioral, and biological data. Previous published work employing spatial analysis of HIV infection has combined two or three of these elements to look at the distributions of HIV and the association with sociodemographic factors (e.g. economic status of a community), physical factors (e.g. roads or healthcare facilities), or co-infections (e.g. Leishmania or malaria) [14–18]. A geographical information system has been used to map samples of drug users, but published work has been limited to describing the distribution of drug use or study participation in cities [19,20]. Geographical proximity to syringe exchange programmes and its association with an increased likelihood of programme use and reduced syringe sharing was the object of one study [21]. Only one study has been published that linked the geographical distribution of HIV prevalence and transmission risk factors [22]. Similarly, studies of the geographical distributions of tuberculosis or sexually transmitted infections have not added both sociodemographic and behavioral variables to the analysis simultaneously [23–26].

Our study demonstrates the power of the spatial approach, because it is only with the addition of geographical data that the relationships among HIV prevalence, frequent and unsafe injection, and incident infections become evident. Whereas a tight geographical clustering of risk and infection might be expected, it has not hitherto been demonstrated. Previous studies have

Fig. 1. Distribution of injection drug users screened at baseline whose residential address within St Petersburg could be determined (N = 788) as a function of the order of recruitment. Recruitment order: 1–160; 161–320; 321–480; 481–640; 641–788.
Fig. 2. Distribution and significant spatial clustering of serological, behavioral, and demographic variables. (a) HIV prevalence; (b) High frequency of heroin injection; (c) Receptive syringe sharing; and (d) Residing with parents.

Table 2. Spatial distribution of demographic and behavioral characteristics of injection drug users in St Petersburg.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Moran’s I ± SD</th>
<th>P value for clustering</th>
<th>Area of clustering, in km² (% of populated area)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV seropositive individuals (N = 236)</td>
<td>0.075 ± 0.030</td>
<td>&lt; 0.01</td>
<td>15.30 (4.1%)</td>
</tr>
<tr>
<td>HIV incidence (N = 18)</td>
<td></td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Age &lt; 24 years (N = 394)</td>
<td>0.117 ± 0.030</td>
<td>&lt; 0.0001</td>
<td>1.10 (0.3%)</td>
</tr>
<tr>
<td>High frequency injectors (N = 307/788)</td>
<td>0.099 ± 0.030</td>
<td>&lt; 0.001</td>
<td>4.56 (1.2%)</td>
</tr>
<tr>
<td>Stimulant injection (N = 299/788)</td>
<td>0.004 ± 0.030</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Receptive syringe sharing (N = 603/788)</td>
<td>0.070 ± 0.030</td>
<td>&lt; 0.05</td>
<td>16.99 (4.6%)</td>
</tr>
<tr>
<td>Sex for drugs or money (N = 68/227)</td>
<td>0.009 ± 0.101</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Employed (N = 351/788)</td>
<td>−0.023 ± 0.030</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Living with parents (N = 502/788)</td>
<td>0.076 ± 0.030</td>
<td>&lt; 0.05</td>
<td>6.58 (1.8%)</td>
</tr>
</tbody>
</table>
shown geographical concentrations of HIV in poor and underserved communities and the dispersion of HIV along roads and highways [27–30]. Not all past studies found such a tight linkage; one study from the San Francisco Bay area using data collected in the late 1990s found that whereas HIV prevalence was indeed concentrated in poor neighborhoods, risky injection practices were, on average, greater among individuals who lived outside of these areas [31]. It was postulated that the emergence of syringe exchange and other harm reduction programmes targeting the poorer communities had resulted in lower risk by the time the study was collected but that prevalence in these communities remained high [32]. Harm reduction was not nearly as widespread or long established in St Petersburg in 2002, when the data for the present report were collected, as it was in the San Francisco Bay area in the late 1990s.

The current study has several limitations. First, the sample of 900 participants, although recruited using three different strategies, must be considered a convenience sample of the estimated 70,000 IDU in St Petersburg. It is thus impossible to determine whether the patterns and associations described in this report are true for the entire population of injectors in the city.

---

**Fig. 3.** Co-clustering of serological, behavioral, and demographic variables. Observed clusters: □ syringe sharing; □ living with parents; ■ heroin use; || prevalence.

**Fig. 4.** Distribution of incident HIV cases within spatial clusters (a) adjacent to the Sitniy market on Petrogradskaya Island and (b) east of the Ladozhskaya metro stop on the east bank of the Neva River. Sitniy and East Neva incidence and prevalence. ○ screened negative; ● prevalent cases; □ incident cases.
Table 3. Comparison of individuals with and without valid St Petersburg residential addresses.

<table>
<thead>
<tr>
<th>Valid addresses (N = 788)</th>
<th>Invalid addresses (N = 112)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV prevalence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.9%</td>
<td>30.4%</td>
<td>NS</td>
</tr>
<tr>
<td>Heroin injection, mean times per week</td>
<td>4.06</td>
<td>4.46</td>
</tr>
<tr>
<td>Syringe sharing, mean times per week</td>
<td>2.36</td>
<td>2.43</td>
</tr>
<tr>
<td>Percentage living with parents</td>
<td>63.3%</td>
<td>61.6%</td>
</tr>
<tr>
<td>Percentage &lt; 24 years old</td>
<td>49.1%</td>
<td>54.5%</td>
</tr>
</tbody>
</table>

Second, much of the data including residential addresses and injection practices are self-reported. Providing an address was a requirement of the screening process and while many of the study participants were later contacted at the addresses they provided, we have no way of knowing what proportion of the participants gave false or misleading addresses. Third, not all of the 900 HPTN033 study participants are included in this analysis. We could not geocode the residential address for 69 participants, and another 53 lived outside the city and were excluded from our analysis. It is possible that the exclusion of these individuals unmasked correlations among those remaining observations. To exclude this possibility we compared those individuals we could (N = 788) and could not (N = 112) geocode to a St Petersburg residential address for all variables found to be significantly clustered (Table 3). There were no differences between these groups when it came to HIV prevalence, injection frequency, syringe sharing, and being younger than 24 years of age. Living with parents was significantly more common among those individuals with a valid St Petersburg residential address.

Despite these limitations, the analysis revealed two important findings. First, the analysis demonstrated the inherent power of adding spatial analysis to traditional biobehavioral explorations of disease transmission. Whereas this idea is not novel, dating back to John Snow’s detection of the link between contaminated water supply and cholera in 1854, new tools for geographical analysis and spatial analysis have simplified the process. The findings have serious implications for the efforts to curtail the HIV epidemic in St Petersburg. With more than 35,000 reported HIV infections city-wide and limited resources, this task might seem insurmountable. Our spatial analysis identified where the epidemic, incident and prevalent cases, were clustered and where behavioral risks such as high injection and syringe sharing frequencies were exacerbating transmission. These locations, covering a small, discrete area of the city, provide a manageable target for prevention programmes such as expanded access to syringe exchange, HIV counseling, testing, and care, and drug abuse treatment.

Authors’ contributions

R.H. participated in the design of participant recruitment and data analysis and led in the writing of the manuscript. R.B. conducted the spatial analysis, prepared the figures, and contributed to the writing of the manuscript. A.V.S. led the collection of data, verified the geographical information, participated in the design of the analysis and in the writing of the manuscript. A.P.K. served as the lead investigator at the study site in St Petersburg and participated in the design of the analysis and the writing of the manuscript. I.F.H. was the principal investigator on the parent HPTN033 study and participated in the design of the analysis and the writing of the manuscript. All authors have seen and approved the final version.

Sponsorship: This work was supported by a grant from the US National Institutes of Health (U01 A147987), as part of the HIV Prevention Trials Network (HPTN), to the University of North Carolina at Chapel Hill and the Biomedical Center, St Petersburg; by the Fogarty International Center (2D43 TW001028) as part of an AIDS International Training and Research Program, which provided training in HIV epidemiology and prevention research to many of the HPTN staff; and by the National Institute on Mental Health (5P30 MH062294) as part of the Yale Center on Interdisciplinary Research on AIDS, which provided support for R.H. and R.B. The US National Institutes of Health provided funding for the research activity, but did not participate in the study design development, collection, analysis or interpretation of data, in report writing or the decision to submit this paper for publication.

Conflicts of interest: None.

References