

Detailed Transmission Network Analysis of a Large Opiate-Driven Outbreak of HIV Infection in the United States

Ellsworth M. Campbell,¹ Hongwei Jia,¹ Anupama Shankar,¹ Debra Hanson,¹ Wei Luo,¹ Silvana Masciotra,¹ S. Michele Owen,¹ Alexandra M. Oster,¹ Romeo R. Galang,¹ Michael W. Spiller,¹ Sara J. Blosser,² Erika Chapman,² Jeremy C. Roseberry,² Jessica Gentry,² Pamela Pontones,² Joan Duwve,^{2,3} Paula Peyrani,⁴ Ron M. Kagan,⁵ Jeannette M. Whitcomb,⁶ Philip J. Peters,¹ Walid Heneine,¹ John T. Brooks,¹ and William M. Switzer¹

¹Division of HIV/AIDS Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia; ²Indiana State Department of Health and ³Indiana University Richard M. Fairbanks School of Public Health, Indianapolis; ⁴Division of Infectious Diseases, University of Louisville, Kentucky; ⁵Quest Diagnostics, Madison, New Jersey; and ⁶LabCorp, Burlington, North Carolina

(See the editorial commentary by Paraskevis on pages 1049–50.)

In January 2015, an outbreak of undiagnosed human immunodeficiency virus (HIV) infections among persons who inject drugs (PWID) was recognized in rural Indiana. By September 2016, 205 persons in this community of approximately 4400 had received a diagnosis of HIV infection. We report results of new approaches to analyzing epidemiologic and laboratory data to understand transmission during this outbreak. HIV genetic distances were calculated using the polymerase region. Networks were generated using data about reported high-risk contacts, viral genetic similarity, and their most parsimonious combinations. Sample collection dates and recency assay results were used to infer dates of infection. Epidemiologic and laboratory data each generated large and dense networks. Integration of these data revealed subgroups with epidemiologic and genetic commonalities, one of which appeared to contain the earliest infections. Predicted infection dates suggest that transmission began in 2011, underwent explosive growth in mid-2014, and slowed after the declaration of a public health emergency. Results from this phylodynamic analysis suggest that the majority of infections had likely already occurred when the investigation began and that early transmission may have been associated with sexual activity and injection drug use. Early and sustained efforts are needed to detect infections and prevent or interrupt rapid transmission within networks of uninfected PWID.

Keywords. HIV outbreak; phylodynamic; transmission network; PWID; transactional sex.

In response to the epidemic of prescription opioid abuse in the United States, pharmaceutical manufacturers have developed novel abuse-deterrent formulations [1]. Crush-resistant abuse-deterrent formulations were designed to discourage insufflation but, according to a recent study [2], may be associated with increased rates of injection. Injection of oxycodone represents a multifaceted public health concern owing to its high potential for overdose and risk of exposure to blood-borne pathogens, such as human immunodeficiency virus (HIV). Persons who inject drugs (PWID) can mitigate these infectious risks by using sterile injection equipment, such as that provided by syringe service programs (SSPs).

However, SSPs are less common in rural than urban or suburban communities and are often explicitly prohibited by law [3].

In 2011, the Indiana State Department of Health investigated a small outbreak of hepatitis C among epidemiologically linked PWID in east central Indiana [4, 5]. In the summer of 2014, nearby Scott County was the epicenter of an unrelated outbreak among PWID that would become one of the largest HIV outbreaks in the United States since the introduction of highly active antiretroviral treatment in the mid-1990s. PWID frequently shared injection equipment while injecting prescription oxycodone, driving the explosive growth of the HIV outbreak [6]. Within this community with high rates of unemployment and low high school graduation rates, exchange of sex for drugs or money (hereafter referred to as transactional sex) was reported by 25% of females infected with HIV who were part of the outbreak, and it likely contributed to the spread of HIV [7, 8].

The original epidemiologic investigation indicated that a single strain of HIV subtype B had spread rapidly within the community of PWID in Scott County [7]. Under conditions of rapid transmission, viral diversity is insufficient to infer specific transmission events from molecular sequence data alone. Even for rapidly evolving pathogens such as HIV, 2 individuals

Received 5 December 2016; editorial decision 22 May 2017; accepted 27 June 2017; published online October 5, 2017.

Presented in part: Conference on Retroviruses and Opportunistic Infections, Boston, MA, 22–25 February 2016, Abstract 215; 23rd International HIV Dynamics and Evolution, Woods Hole, MA, 24–27 April 2016, Abstract 24; Public Health Informatics Conference, Atlanta, GA, 21–24 August 2016, Abstract D11.

Correspondence: E. M. Campbell, MSc, Centers for Disease Control and Prevention, Division of HIV/AIDS Prevention, 1600 Clifton Rd, Bldg 18, Fl 2, Rm 113, Atlanta, GA 30329 (ells@cdc.gov).

The Journal of Infectious Diseases® 2017;216:1053–62

Published by Oxford University Press for the Infectious Diseases Society of America 2017. This work is written by (a) US Government employee(s) and is in the public domain in the US. DOI: 10.1093/infdis/jix307

infected by genetically identical viruses may never have been in direct contact [9, 10]. The likelihood that 2 infections represent a transmission pair can be augmented by a more holistic analysis of relevant data sources. The emerging field of phylodynamics seeks to understand evolutionary processes through the integration of epidemiologic data (eg, behavioral risk factors and high-risk contacts identified through contact tracing) with biologic data (eg, immunologic response and pathogen genetic sequences). One goal of a phylodynamic analysis is to obtain an improved picture of the historical sequence of transmissions [11–14], represented as a network that, unlike a phylogeny, is directly comparable to contact networks. Phylodynamic methods have been used to define inclusion and exclusion criteria for an outbreak [15], prioritize resource allocation during tuberculosis outbreak investigations [16], assess the effectiveness of intervention and prevention efforts [17–19], and predict [16] and identify [15] venues where public health interventions can be applied.

Transmission network analysis can provide public health officials with high-resolution information about transmission and drug resistance dynamics at global [20], national [21], and subnational [22, 23] levels. These data have been used to identify growing transmission clusters of concern, monitor transmission of antiretroviral drug resistance, prioritize high-risk groups for prevention efforts, and aid epidemiologic investigations [24–27]. While transmission network analysis is a well-established complement to tuberculosis contact-tracing investigations [16, 28, 29], it is a relatively new approach for HIV surveillance and outbreak investigation. We conducted phylodynamic analysis of epidemiologic, phylogenetic, and clinical laboratory data to infer and characterize the structure and growth rate of the recent HIV outbreak in rural Indiana.

METHODS

The Centers for Disease Control and Prevention (CDC) generated HIV polymerase (*pol*) Sanger sequences from serum/plasma samples or received sequences from commercial laboratories from persons with a diagnosis of HIV infection who met the case definition for inclusion in the outbreak [6]. We used BLAST to compare *pol* sequences from these case patients against >1.2 million HIV sequences from GenBank, commercial databases, and the CDC's Molecular HIV Surveillance database [30] to determine whether the outbreak sequence was closely related to any other sequences and, if so, the geographic location where those diagnoses occurred. Sequences were used to construct a phylogenetic tree by an approximately maximum-likelihood method (Supplementary Materials). Serum and plasma specimens also underwent Bio-Rad avidity incidence (BRAI) testing to estimate the recency of infection (Supplementary Materials). Viral loads for HIV-infected cases and *pol* sequences for some cases were provided by 2 commercial laboratories (LabCorp and Quest Diagnostics). The Indiana Communicable Disease

Reporting Rule requires that HIV-related data be reported by medical laboratories weekly to the Indiana State Department of Health. All HIV-positive case patients were interviewed by using a standardized form to ascertain sex and injection drug-using partners. To identify all potential at-risk persons within the community, individuals were also asked to name anyone who they believed might benefit from HIV testing.

Behavioral Contribution of Risk

The total number of high-risk sex, needle sharing, and both sex and needle sharing contacts a person reports can serve as a proxy measure of their exposure risk, regardless of that individual's HIV status. These data can be represented as a contact network in which individuals are nodes that are connected by lines corresponding to the type of contact. We also included each individual's number of unique high-risk contacts, age, and sex as input variables to construct decision trees, which constitute a predictive machine learning method whose outcome variable was HIV status. A forest of decision trees was trained iteratively, using a random sample of 25% of these data. The remaining 75% were used to assess the sensitivity and specificity of the model classification. Decision tree analysis was used to identify simple demographic and behavior-based rules that were predictive of population-level HIV infection in this outbreak that could also inform future outbreak prevention, intervention, and investigation decision-making.

Genetic Distance Network

Genetic distance (d) was determined for all pairs of HIV *pol* sequences after codon alignment of each sequence to HIV_HXB2 (GenBank accession number K03455), using a nucleotide substitution model [20]. Genetic distance networks were constructed according to distances between *pol* sequences, wherein each sequence was linked to other closely related sequences [20]. We include distances between outbreak-associated sequences and references obtained from Indiana and surrounding states between 2004 and 2015. Previously, a distance threshold of 1.5% was used to infer transmission, based on the finding of approximately 1% intrahost HIV subtype B divergence in persons from the United States over a decade of infection [20]. Reliance on a single genetic distance threshold may artificially discard relevant genetic links. To incorporate both early and late transmissions, we constructed a minimum spanning tree (MST), which, similar to a phylogeny, is the minimum number of genetic links that still maintain connections between all nodes. An MST can be seen as the most parsimonious subset of genetic distances required to produce the complete network. There can be many equivalent MSTs for a given network, especially when distances are uniform. To account for this range of possibilities, we constructed many nonisomorphic (unique) MSTs for all genetic links with $\leq 1.5\%$ distance, wherein edges were weighted by their frequency of occurrence across all MSTs.

Inferred Transmission Network

We considered a report of high-risk contact between 2 HIV-positive persons as a potential transmission event when the link corresponded to one found among the forest of MSTs. When an individual did not report a high-risk contact that corresponded to a well-supported genetic linkage found among the MSTs, their HIV *pol* sequence was linked to its closest genetic neighbor, and that link was designated an unreported putative transmission event. If there were multiple equidistant close genetic neighbors, those who appeared among all MSTs more than once were also included. If the closest genetic neighbor was another individual with no high-risk contacts, we included the link to the closest genetic neighbor who was already a member of the largest connected component.

Transmission Cluster Growth

Serological determination of recent HIV infection by measuring antibody binding strength or avidity to HIV antigens has been used to measure incidence, understand transmission dynamics, and evaluate prevention strategies. One validated method for determining recent infection is the CDC-modified Bio-Rad HIV enzyme immunoassay (BRAI) that calculates an avidity index (Supplementary Materials) [31]. We developed a novel bioinformatics method to use avidity index scores to infer possible HIV infection dates. First, a *k*-means clustering method was used to define the distribution of 3 distinct clusters of recency from the outbreak. The avidity index minima and maxima of each cluster of recency were then used to create subsets of corresponding observations from separate longitudinal seroconversion cohorts of individuals [31–33] for whom the inferred date of detectable infection was the midpoint between their dates of last negative and first positive HIV test result. The duration of recent infection for individuals included in the seroconversion panel was defined as the difference between the sample collection date and the date of the last negative test result. For each outbreak-associated individual, a duration of recent infection was assigned a randomly selected duration from the seroconversion cohort according to their respective recency cluster. An individual's inferred date of infection is estimated by subtracting the simulated duration of recent infection from the specimen collection date. To directly compare the predicted incidence curve with the observed diagnosis curve, persons with inconclusive or missing avidity index results were randomly assigned a duration of recent infection from the entire seroconversion cohort. To account for stochastic variation, these processes were repeated 1000 times to infer a best-fit epidemiological curve and its computed standard error. We used the R package *EpiEstim* [34] to infer the instantaneous reproduction number (ie, transmissibility at a certain time step) over a 30-day sliding window beginning in late 2014, under the assumption of a serial interval that is normally distributed with a mean and standard deviation of 1 week. Finally, we estimated

the reproduction number (R_0) of the outbreak by assuming that $R_0 = \tau \cdot \bar{c} \cdot d$, where τ is the mean number of high-risk contacts infected with HIV divided by the mean number high-risk contacts, \bar{c} is the mean number of high-risk contacts, and d is the total duration of the outbreak [35].

RESULTS

Behavioral Contribution of Risk

Contact tracing yielded a network of 1060 high-risk linkages among 411 individuals, including 183 individuals who received a diagnosis of HIV infection during the outbreak investigation through 3 March 2016 (Supplementary Figure 1). Of these reported high-risk contacts, 79.2% (839) were injection drug users only, 7.9% (84) were sex partners only, and 12.9% (137) were sex partners and injection drug users (Figure 1). During the investigation, 1 person infected with HIV was found to have received a diagnosis of HIV infection nearly a decade prior; all other diagnoses occurred during or after November 2014.

The prevalence of self-reported injection drug use was 91.8% (168 of 183) among HIV-positive individuals and 86.6% (356 of 411) among all persons in the high-risk contact network. Persons who “would benefit from testing” (313 individuals) were the second most common contact type reported during the investigation. HIV-positive individuals were named as someone who “would benefit from testing” by persons with whom they did not have a high-risk contact at a rate 2.3 times that for those who tested negative for HIV ($P < .01$). When interviewed, 92.3% (60 of 65) who were identified as someone who “would benefit from testing” reported injection drug use in the prior year. The

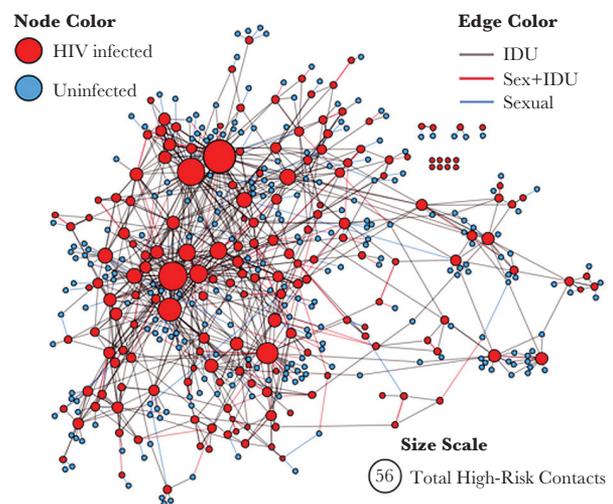


Figure 1. Contact tracing network, human immunodeficiency virus (HIV) outbreak, Scott County, Indiana, 2015. Each blue (uninfected) circle represents a person at high risk of HIV infection. Red circles represent HIV-positive individuals. Each circle is sized according to the number of high-risk contacts reported by the corresponding individual or partner, with scale shown in the figure. IDU, injection drug use contact; sex, sexual contact; sex+IDU, sexual and injection drug use contact.

person with the earliest HIV diagnosis reported no high-risk contacts or injection drug use. However, this individual was identified to an outbreak investigator as someone who “would benefit from testing” by an individual who reported transactional sex and 34 unique high-risk contacts; additionally, each person identified by this individual as someone who “would benefit from testing” eventually received a diagnosis of HIV infection. The distribution of injection drug-using contacts among HIV-infected persons (mean number [\pm SD], 4.1 ± 6.2) had a long tail owing to the 24.6% of persons (33 of 134) who reported a number of injection drug-using contacts that was at least 1 SD above the mean (ie, >10 ; Figure 2A).

Decision tree analysis yielded 5 terminal nodes in which individuals were grouped according to the number of high-risk contacts (Figure 3). Of 117 individuals who reported >3 injection drug-using partners, 93.2% (109) were infected with HIV. Individuals with ≤ 3 injection drug-using partners were at much lower risk of HIV infection (16.8%; 73 of 435). However, 48.7% of individuals (38 of 78) with ≤ 3 injection drug-using partners and at least 1 sex and injection drug-using partner had HIV infection. Finally, 87.1% of individuals (27 of 31) who reported 2–3 injection drug-using partners and at least 1 sex and injection drug-using partner had HIV infection.

Genetic Distance Network

Under a standard genetic distance threshold ($d \leq 1.5\%$) that was meant to capture potential transmission events in the last 10 years [20, 36], the *pol* network consisted of 14877 densely connected potential transmission pairs, resulting in a single, highly connected network component. Among these potential pairs, 3327 (22.4%) involved extremely close genetic linkages that represented approximately 1–2 base substitution(s) between sequences, based on a lower genetic distance threshold ($d \leq 0.1\%$; Figure 4A), which confirmed that a single HIV strain was associated with the outbreak. Application of this lower threshold revealed community structure within the largest connected component ($N_A = 72$, $N_B = 23$, and $N_C = 24$) that corresponded to 3 subgroups identified in the phylogenetic tree (Figure 4B). While application of this strict threshold increased the resolution of the community structure, it also caused singletons and small clusters (size ≤ 5) to dissociate from the outbreak cluster.

Rather than retaining only genetic distances that fell below a threshold (Figure 5A), we also considered a forest of unique MSTs of all transmission pairs in which $d \leq 1.5\%$. The composite of unique MSTs was strikingly similar to the $\leq 0.1\%$ genetic distance network, with a few notable exceptions (Figure 5B). Perhaps most noticeable was the inclusion of the aforementioned singletons and small clusters into the 3 previously defined subgroups ($N_A = 75$, $N_B = 25$, and $N_C = 26$) and the generation of 3 new subgroups ($N_D = 7$, $N_E = 4$, and $N_F = 3$). The *pol* sequence from the case diagnosed a

decade prior clustered among an outlier subgroup ($N_O = 10$ sequences) that were most closely related to sequences in subgroup A, as well as the sequences identified to bridge subgroup A to subgroups B and C (Figure 5). Using a multinomial regression model (Supplementary Materials), members of subgroup B were more likely to report ≥ 2 concurrent sex partners ($P < .10$; Supplementary Table 1), relative to members of subgroup C. Subgroup O sequences fell between reference sequences obtained in Indiana between 2004 and 2015 and all other sequences (Figure 5). Relative to subgroup C, members of subgroup O were more likely to report transactional sex ($P < .05$) and >2 sexual contacts ($P < .10$) and more likely to report no injection drug use contacts ($P < .10$). Members of subgroup O were also more likely to be classified by recency testing as established infections ($P < .10$; Supplementary Table 1).

Inferred Transmission Network

The inferred transmission network, constructed from both the contact tracing and MST networks, consisted of 176 nodes connected by 303 potential transmission events, 52.3% (159) of which involved injection drug use only, 6.3% (19) involved sex and injection drug use, and 0.9% (3) involved sex only (Figure 5C); 40.6% of potential transmission events (123) did not have a corresponding epidemiological link and were designated unreported. The mean genetic distance between *pol* sequences of unreported links were differentiated by only approximately 1–2 mutations ($d \leq 0.1\%$). The majority of high-risk contacts (66.0%; 460 of 697) between HIV-infected individuals were not supported by a close genetic linkage found among all MSTs, as only 34% of high-risk partners were determined to be potential transmission partners. Subgroups B–F could be characterized by a single bridging sequence that separated their members from subgroup A. Subgroup O, which contained the sequence from the earliest known HIV diagnosis received by a case patient, was not differentiated from subgroup A because there were 8 potential transmission links between subgroups O and A, of which 3 were designated unreported.

Inferred Dates of Infection

BRAI were separated into 3 categories (center points, 8.0, 60.4, and 97.6) by *k*-means clustering analysis. Using these 3 categories, we randomly drew dates of seroconversion linked empirically to BRAI values in an iterative process to derive the incidence curve for the outbreak. Predicted transmission events began in 2011 and reached an exponential growth phase by mid-2014 (Figure 6). By the date of the first diagnosis in late 2014, 41% of persons (75 of 183) who would receive a diagnosis of HIV infection in 2015 were already seropositive. By 26 March 2015, the observed inflection point of the diagnosis curve and the date that the governor of Indiana declared a state public health emergency, $>80\%$ of the HIV infections

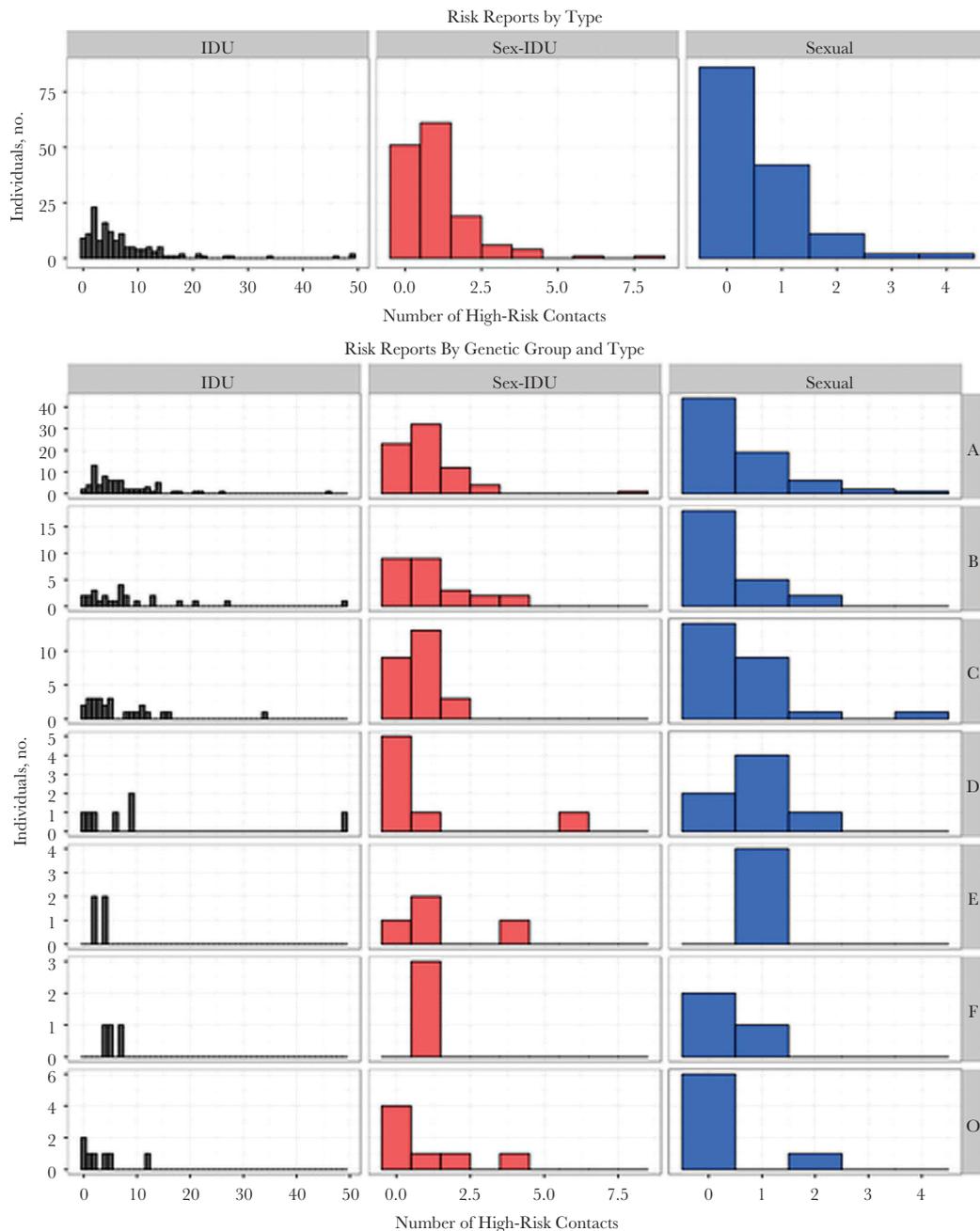


Figure 2. Distribution of risk exposure type by the total number of unique reports of high-risk contact per individual, by contact type (injection drug use contact [IDU], sexual contact only [sex], and both sexual and injection drug use contact [sex+IDU]; A) and the total number of unique reports of high-risk contact per individual by contact type and viral genetic subgroup (A–F and O; B). Note scaling differences on x-axes for each contact type.

that would eventually be diagnosed and included in this analysis had already occurred. The estimated R_0 was 3.8, suggesting that a reduction of overall transmissibility by 75% is required to prevent an outbreak of this scale. A separate analysis of the outbreak's growth potential illustrated that the instantaneous reproduction number fell below the epidemic threshold (ie, it was <1 ; Supplementary Figure 2), just 3 weeks after the state public health emergency was declared and as the syringe exchange program launched. This finding suggests that awareness

of the outbreak, initiation of SSPs, and availability of additional healthcare resources may have curtailed new infections.

DISCUSSION

Pathogen sequence data are traditionally represented as a phylogenetic tree, but this approach does not integrate potentially useful epidemiologic information about high-risk contacts of case patients that are gathered routinely during outbreak investigations. By representing genetic differences between viruses

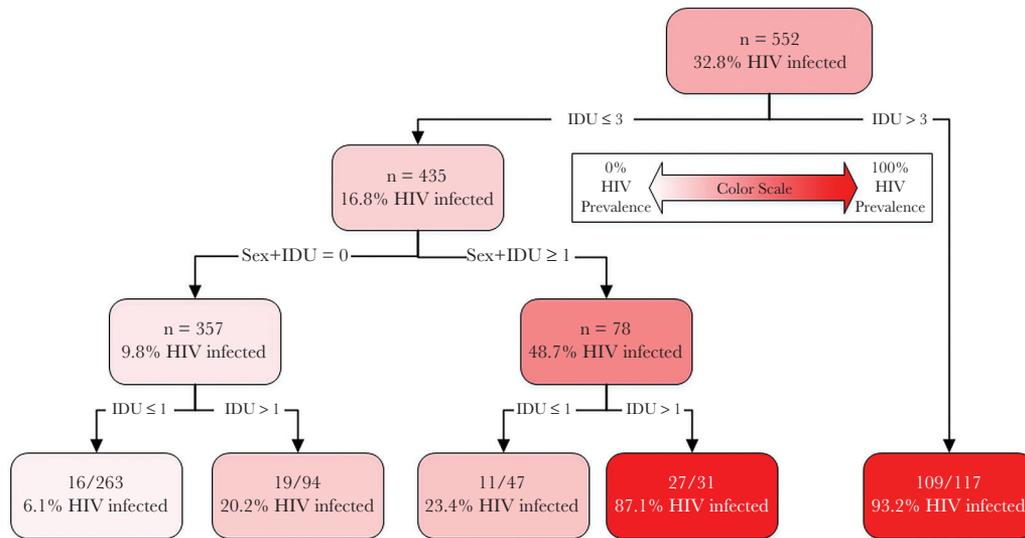


Figure 3. Quantification of human immunodeficiency virus (HIV) infection risk levels, by decision tree analysis based on behavioral risk factors, including number of injection drug–using and sex partners. The decision tree is colored and scaled with respect to the prevalence of HIV-positive persons within each node. Each split in the decision tree is statistically significant ($P < .05$). IDU, injection drug use contacts; sex, sexual contacts; sex+IDU, sexual and injection drug use contacts.

as a network, these data can be directly mapped onto epidemiologic evidence. We found that, under conditions of recent and rapid HIV transmission, as occurred in this outbreak and as evidenced by minimal viral diversity, incidence testing, and outbreak simulations, a relaxed genetic distance threshold of 1.5% is too inclusive to discern specific transmission events.

However, reduction of this inclusion threshold to 0.1% unnecessarily sacrificed links that may represent the earliest infections that were phylogenetically and epidemiologically linked to the outbreak. To overcome this limitation, we characterized transmission patterns by comparing only the most parsimonious genetic links among high-risk contacts. Owing

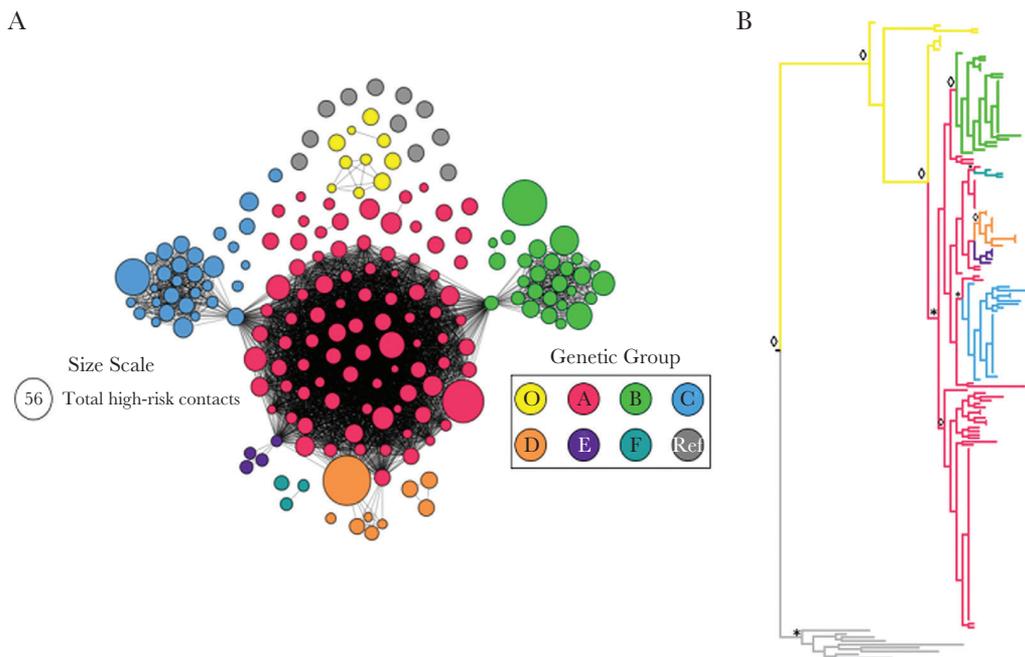


Figure 4. Human immunodeficiency virus (HIV) polymerase (*pol*) sequence analyses. Genetic subgroups in the phylogeny (branches) and network (nodes) are colored as shown in the key. *A*, Phylogeny of 183 HIV *pol* sequences, colored by genetic subgroup, from the Indiana outbreak and local Indiana reference *pol* sequences (gray). Phylogenetic analyses were conducted using FastTree maximum likelihood analysis. Circles are sized according to the corresponding individual's number of reported high-risk contacts. Confidence values for branching patterns were assessed by using the Shimodaira-Hasegawa (sh) test and are given as probabilities at nodes (*, $sh > 0.80$; \diamond , $sh > 0.90$). *B*, Genetic distance (d) network, using a d threshold of $\leq 0.1\%$. Ref, reference.

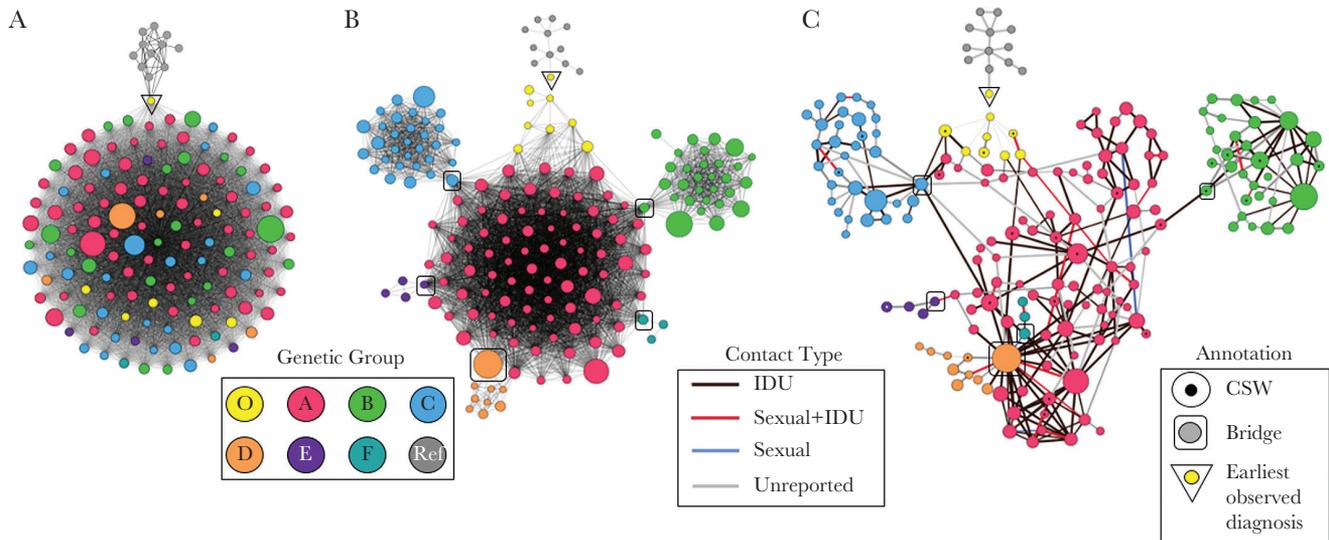


Figure 5. Genetic distance and inferred human immunodeficiency virus (HIV) transmission networks. Circles represent a polymerase (*pol*) sequence isolated from an HIV-infected person. Lines represent close genetic links between HIV sequences. Genetic subgroups in the network (nodes) are colored as shown in the key. *A*, HIV *pol* genetic distance (d) network, using a d cutoff of $\leq 1.5\%$. *B*, Nonisomorphic minimum spanning trees (MSTs) of the HIV *pol* genetic distance network. *C*, Inferred transmission network based on the synthesis of the *pol* genetic distance network in panel *A* and the MSTs in panel *B*. Lines represent inferred transmission events, with the thickness proportional to $1.0 - d$. Bridge, sequence from person in network linking genetic subgroups; CSW, commercial sex worker contact and/or transactional sexual contact; earliest observed diagnosis, person in network identified during outbreak investigation who had the earliest HIV infection diagnosis; IDU, injection drug use; unreported, contact information not reported.

to inherent biases in self-reported data about socially sensitive topics (ie, sexual behavior and illegal injection of drugs), the existence or absence of any link cannot be confirmed. The distinct subgroups we identified in the genetic and transmission networks were likely the result of a combination of geographic, social, and temporal factors. The presence of discrete subgroups suggests that the prevention of explosive HIV transmission is difficult but that, if detected sufficiently early, there may be subsequent transmission bottlenecks where

application of targeted preventive strategies, such as SSPs, can interrupt transmission.

Integration of epidemiologic and laboratory data in a phylogenetic model facilitated a deeper understanding of HIV transmission dynamics in an outbreak among PWID in rural America, where an unprecedented epidemic of opioid and heroin abuse and injection drug use is growing. We demonstrated that reports of high-risk contact collected by disease investigation specialists can be integrated with pathogen sequence data to efficiently focus

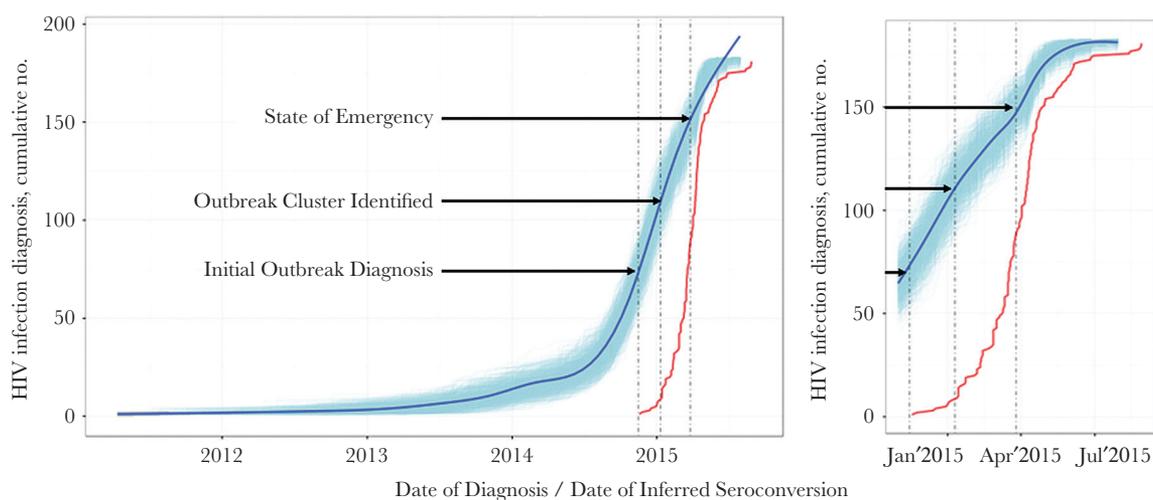


Figure 6. Cumulative human immunodeficiency virus (HIV) diagnoses (red) and simulated incidence (light blue), by date. Mean cumulative seroconversion dates over time are in dark blue. Dot-dashed vertical gray lines indicate important dates during the outbreak response and their corresponding locations on the curve (horizontal black arrows). By use of avidity indices determined by serological HIV recency testing (Supplementary Materials), outbreaks were simulated to infer the time course of the outbreak. The panel on right shows an expanded view of results on the timeline during the first half of 2015.

contact-tracing efforts on putative transmission events. We used a novel bioinformatics method that uses biomarker incidence results (ie, HIV recency testing) to quantify both the transmission cluster's growth rate and infer the potential period in which the outbreak began. We found that transmission of a single geographically localized strain of HIV may have begun as early as 2011 and that the majority of infections had likely already occurred prior to the declaration of a state public health emergency. We also determined that persons infected with HIV in one genetic subgroup were among the earliest individuals infected during the outbreak. These same individuals were characterized as more likely to engage in high-risk sexual behavior than to use injection drugs. When coupled with the lack of self-reported injection drug use behavior by the person with the earliest diagnosis of HIV infection, these findings suggest that high-risk sexual contact was the most plausible route by which HIV was introduced into this population of PWID. Subgroups in the transmission network were largely similar with respect to demographic composition and high-risk behaviors, although one subgroup reported more concurrent high-risk partners. Taken together, these results demonstrate that an integrated phylodynamic approach, if applied in a timely fashion, can provide deeper information about the spread of infection, including both the historical and future trajectories of an outbreak that may allow prevention strategies to be tailored more rapidly and specifically to a community's needs. Automation of such analyses could provide actionable and timely intelligence to aid decision making by public health experts in the field.

In summary, we have demonstrated the utility of combining epidemiologic and laboratory data in a phylodynamic analysis to characterize an outbreak of HIV infections in a rural community of PWIDs that was initiated by a single strain of HIV. Our findings and methods may benefit future outbreak investigations by identifying, in near real-time and interactive fashion, subgroups or individuals at greatest risk for onward transmission of the pathogen and potential bridge individuals who could be prioritized for intervention as the outbreak evolves. We have also described a novel bioinformatic method based on a recency assay used to infer dates of HIV seroconversion. With this approach, we showed that the majority of HIV infections in the outbreak likely occurred prior to the declaration of a state public health emergency that included an SSP for affected counties. Had an SSP been in place prior to recognition of the outbreak, the explosive phase of the outbreak may have been blunted. Evolving changes in federal and state laws for SSPs will improve efforts to prevent similar outbreaks in the future [37]. As the heroin and opioid epidemic continues to worsen in the United States, it is imperative that susceptible populations are protected and outbreaks are quickly identified and intervened upon to prevent resurgence and rapid growth of injection drug use-associated HIV infections. Initiation or expansion of test-and-treat strategies for HIV and hepatitis virus infections in these opioid-epidemic areas affecting mostly rural populations not previously considered vulnerable to these conditions may expedite prevention efforts.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. We thank the study participants and their partners and the Indiana, Scott County, and Clark County departments of health, for their rigorous public health efforts during and after the outbreak investigation. We also thank Leidos Innovations, formerly Lockheed Martin IS&GS, for assistance with machine learning applications.

Disclaimer. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. The use of trade names and commercial sources is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention.

Financial support. This work was supported by the state government of Indiana, the Scott and Clark County health departments, and the Department of Health and Human Services.

Potential conflicts of interest. Dr Muzny reports grants from NIH/NIAID outside the submitted work. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Alexander L, Mannion RO, Weingarten B, Fanelli RJ, Stiles GL. Development and impact of prescription opioid abuse deterrent formulation technologies. *Drug Alcohol Depend* **2014**; 138:1–6.
2. Cicero TJ, Ellis MS, Kasper ZA. A tale of 2 ADFs: differences in the effectiveness of abuse-deterrent formulations of oxycodone and oxycodone extended-release drugs. *Pain* **2016**; 157:1232–8.
3. Des Jarlais DC, Nugent A, Solberg A, Feelemyer J, Mermin J, Holtzman D. Syringe service programs for persons who inject drugs in urban, suburban, and rural areas—United States, 2013. *MMWR Morb Mortal Wkly Rep* **2015**; 64:1337–41.
4. Strathdee SA, Beyrer C. Threading the needle—how to stop the HIV outbreak in rural Indiana. *N Engl J Med* **2015**; 373:397–9.
5. McFarlane T. Emergence of acute Hepatitis C in young Indiana residents who inject drugs [Internet]. Indianapolis, IN: Indiana State Department of Health, **2011**. http://www.in.gov/isdh/files/HCV_Cluster_Investigation_-_Final.pdf. Accessed 1 December 2016.

6. Conrad C, Bradley HM, Broz D, et al.; Centers for Disease Control and Prevention (CDC). Community outbreak of HIV infection linked to injection drug use of oxymorphone—Indiana, 2015. *MMWR Morb Mortal Wkly Rep* **2015**; 64:443–4.
7. Peters PJ, Pontones P, Hoover KW, et al. An outbreak of HIV infection linked to injection drug use of oxymorphone—Indiana, 2014–2015. *N Engl J Med* **2016**.
8. Van Handel MM, Rose CE, Hallisey EJ, et al. County-level vulnerability assessment for rapid dissemination of HIV or HCV infections among persons who inject drugs, United States. *J Acquir Immune Defic Syndr* **2016**; 73:323–31.
9. Romero-Severson EO, Bulla I, Leitner T. Phylogenetically resolving epidemiologic linkage. *Proc Natl Acad Sci U S A* **2016**; 113:2690–5.
10. Frost SD, Pillay D. Understanding drivers of phylogenetic clustering in molecular epidemiological studies of HIV. *J Infect Dis* **2015**; 211:856–8.
11. Holmes EC, Grenfell BT. Discovering the phylodynamics of RNA viruses. *PLoS Comput Biol* **2009**; 5:e1000505.
12. Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **2004**; 303:327–32.
13. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* **2008**; 5:e50.
14. Paraskevis D, Nikolopoulos G, Fotiou A, et al. Economic recession and emergence of an HIV-1 outbreak among drug injectors in Athens metropolitan area: a longitudinal study. *PLoS One* **2013**; 8:e78941.
15. De P, Singh AE, Wong T, Yacoub W, Jolly AM. Sexual network analysis of a gonorrhoea outbreak. *Sex Transm Infect* **2004**; 80:280–5.
16. Cook VJ, Sun SJ, Tapia J, et al. Transmission network analysis in tuberculosis contact investigations. *J Infect Dis* **2007**; 196:1517–27.
17. Robinson SE, Everett MG, Christley RM. Recent network evolution increases the potential for large epidemics in the British cattle population. *J R Soc Interface* **2007**; 4:669–74.
18. Little SJ, Kosakovsky Pond SL, Anderson CM, et al. Using HIV networks to inform real time prevention interventions. *PLoS One* **2014**; 9:e98443.
19. Wertheim JO, Kosakovsky Pond SL, Little SJ, De Gruttola V. Using HIV transmission networks to investigate community effects in HIV prevention trials. *PLoS One* **2011**; 6:e27775.
20. Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. *J Infect Dis* **2014**; 209:304–13.
21. Aldous JL, Pond SK, Poon A, et al. Characterizing HIV transmission networks across the United States. *Clin Infect Dis* **2012**; 55:1135–43.
22. Poon AF, Joy JB, Woods CK, et al. The impact of clinical, demographic and risk factors on rates of HIV transmission: a population-based phylogenetic analysis in British Columbia, Canada. *J Infect Dis* **2015**; 211:926–35.
23. Oster AM, Pieniazek D, Zhang X, et al. Demographic but not geographic insularity in HIV transmission among young black MSM. *AIDS* **2011**; 25:2157–65.
24. Oster AM, Wertheim JO, Hernandez AL, Ocfemia MC, Saduvala N, Hall HI. Using molecular HIV surveillance data to understand transmission between subpopulations in the United States. *J Acquir Immune Defic Syndr* **2015**; 70:444–51.
25. Valverde EE, DiNenno EA, Schulden JD, Oster A, Painter T. Sexually transmitted infection diagnoses among Hispanic immigrant and migrant men who have sex with men in the United States. *Int J STD AIDS* **2016**; 27:1162–9.
26. Whiteside YO, Song R, Wertheim JO, Oster AM. Molecular analysis allows inference into HIV transmission among young men who have sex with men in the United States. *AIDS* **2015**; 29:2517–22.
27. Epidemic Intelligence Service. Epi-Aid—epidemiologic assistance fact sheet [Internet]. Atlanta, GA: Centers for Disease Control and Prevention, **2015**. <http://www.cdc.gov/eis/downloads/epi-aid-fact-sheet.pdf>. Accessed 1 December 2016.
28. Andre M, Ijaz K, Tillinghast JD, et al. Transmission network analysis to complement routine tuberculosis contact investigations. *Am J Public Health* **2007**; 97:470–7.
29. Walker TM, Monk P, Smith EG, Peto TE. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing. *Clin Microbiol Infect* **2013**; 19:796–802.
30. Wertheim JO, Oster AM, Hernandez AL, Saduvala N, Bañez Ocfemia MC, Hall HI. The international dimension of the U.S. HIV transmission network and onward transmission of HIV recently imported into the United States. *AIDS Res Hum Retroviruses* **2016**; 32:1046–53.
31. Hanson DL, Song R, Masciotra S, et al. Mean recency period for estimation of HIV-1 incidence with the BED-capture EIA and Bio-Rad avidity in persons diagnosed in the United States with subtype B. *PLoS One* **2016**; 11:e0152327.
32. Flys T, Nissley DV, Claassen CW, et al. Sensitive drug-resistance assays reveal long-term persistence of HIV-1 variants with the K103N nevirapine (NVP) resistance mutation in some women and infants after the administration of single-dose NVP: HIVNET 012. *J Infect Dis* **2005**; 192:24–9.
33. Koblin BA, Heagerty P, Sheon A, et al. Readiness of high-risk populations in the HIV Network for Prevention Trials to participate in HIV vaccine efficacy trials in the United States. *AIDS* **1998**; 12:785–93.

34. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* **2013**; 178:1505–12.
35. May RM, Gupta S, McLean AR. Infectious disease dynamics: What characterizes a successful invader? *Philos Trans R Soc Lond B Biol Sci* **2001**; 356:901–10.
36. Bernard EJ, Azad Y, Vandamme AM, Weait M, Geretti AM. HIV forensics: pitfalls and acceptable standards in the use of phylogenetic analysis as evidence in criminal investigations of HIV transmission. *HIV Med* **2007**; 8:382–7.
37. US Congress. The Fiscal Year 2016 Omnibus Appropriations Act. Section 520, H.R. 2029. 114th Congress, **2016**.