



# HHS Public Access

Author manuscript

*J Acquir Immune Defic Syndr.* Author manuscript; available in PMC 2025 April 01.

Published in final edited form as:

*J Acquir Immune Defic Syndr.* 2024 April 01; 95(4): 355–361. doi:10.1097/QAI.0000000000003367.

## Simulation of full HIV cluster networks in a nationally representative model indicates intervention opportunities

Anne Marie France, PhD, MPH<sup>1</sup>, Nivedha Panneer, MPH<sup>1</sup>, Paul G. Farnham, PhD<sup>1</sup>,  
Alexandra M. Oster, MD<sup>1</sup>, Alex Viguerie, PhD<sup>1</sup>, Chaitra Gopalappa, PhD<sup>1,2</sup>

<sup>1</sup>Division of HIV Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention

<sup>2</sup>University of Massachusetts Amherst, Amherst, MA, United States

### Abstract

**Background:** Clusters of rapid HIV transmission in the United States are increasingly recognized through analysis of HIV molecular sequence data reported to the National HIV Surveillance System. Understanding the full extent of cluster networks is important to assess intervention opportunities. However, full cluster networks include undiagnosed and other infections that cannot be systematically observed in real life.

**Methods:** We replicated HIV molecular cluster networks during 2015–2017 in the United States using a stochastic dynamic network simulation model of sexual transmission of HIV. Clusters were defined at the 0.5% genetic distance threshold. Ongoing priority clusters had growth of 3 diagnoses/year in multiple years; new priority clusters first had 3 diagnoses/year in 2017. We assessed the full extent, composition and transmission rates of new and ongoing priority clusters.

**Results:** Full clusters were 3–9 times larger than detected clusters, with median detected cluster sizes in new and ongoing priority clusters of 4 (range 3–9) and 11 (range 3–33), respectively, corresponding to full cluster sizes with a median of 14 (3–74) and 94 (7–318), respectively. A median of 36.3% (range 11.1–72.6%) of infections in the full new priority clusters were undiagnosed. HIV transmission rates in these clusters were >4 times the overall rate observed in the entire simulation.

**Conclusions:** Priority clusters reflect networks with rapid HIV transmission. The substantially larger full extent of these clusters, high proportion of undiagnosed infections and high transmission rates indicate opportunities for public health intervention and impact.

### Keywords

HIV molecular epidemiology; HIV transmission cluster; Network simulation model; Transmission rate; undiagnosed infections

---

Corresponding author: Anne Marie France 1600 Clifton Rd., NE, Mailstop H24-5, Atlanta, GA 30329, (o) 404-639-5313, (f) 404-471-8503, hgj6@cdc.gov.

**Disclaimer:** The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

The authors have no conflicts of interests to disclose.

## Introduction

Responding to HIV clusters and outbreaks (i.e., cluster detection and response [CDR]) is one of four pillars of the U.S. federal *Ending the HIV Epidemic in the U.S. (EHE)* initiative. HIV clusters or outbreaks are groups of people that who are experiencing rapid HIV transmission. Cluster detection identifies this rapid transmission to guide public health interventions, identifying gaps where care and prevention services are not effectively reaching networks in which transmission is occurring<sup>1</sup>. While multiple approaches can detect clusters and outbreaks, large HIV transmission clusters are increasingly recognized through analysis of HIV molecular sequence data reported to the National HIV Surveillance System (NHSS)<sup>2</sup>. Current Centers for Disease Control and Prevention (CDC) molecular cluster detection methods focus on recent, rapid transmission.

HIV mutates rapidly, and molecular cluster detection methods distinguish infections that are more closely related in transmission networks from more distantly related infections. The amount of difference between two viruses' nucleotide sequences ('genetic distance') indicates how closely they are related; viruses with very similar sequences are more closely related and have a smaller genetic distance, while those with more differences are more distantly related and have a larger genetic distance. Clusters can be defined by applying a threshold of genetic distance; any sequence within a given genetic distance of at least one other sequence is considered part of a cluster<sup>3,4</sup>. Analyses have found that clusters meeting CDC's national priority cluster definition, based on a 0.5% genetic distance threshold<sup>5</sup>, reflect rapid HIV transmission<sup>2,5,6</sup>, with transmission rates >8x the overall U.S. HIV transmission rate<sup>5</sup>.

Only infections which have been diagnosed, with a sequence reported to public health surveillance, can be detected as part of a molecular cluster. Epidemiologic investigation often reveals substantially larger transmission networks underlying detected clusters, with evidence that infections were acquired recently<sup>1,7</sup>. Knowledge of the full extent and composition of detected clusters is limited, as the full cluster cannot be systematically observed in real life, even with the most robust of epidemiologic investigations (i.e., contact tracing). Understanding the full extent of clusters, including undiagnosed infections, is important to assess opportunities for intervention and potential impact. Moreover, estimates of cluster-level metrics (e.g., transmission rates) when considering the full cluster/network can provide valuable information for cluster prioritization.

A simulation model can be used to assess the potential impact of CDR activities, including impacts of specific intervention strategies, and methods for HIV CDR. Agent-based simulation models are well suited to understand the complex, interdependent systems that underlie molecular clusters of pathogens<sup>8</sup>. We implemented an algorithm to detect molecular clusters in a nationally representative, stochastic, agent-based network model of HIV transmission. The algorithm uses time as a proxy for viral evolution, allowing the identification of molecular clusters across a range of genetic distance thresholds.

This model was previously validated against overall population-level metrics of HIV transmission in the United States, including prevalence, incidence, and care continuum, as

well as clustering at the 0.5% genetic distance threshold<sup>9</sup>. To ensure the model replicated the complex dynamics generating transmission clusters, in this analysis we include an expanded validation comparing cluster distributions across multiple genetic distance thresholds and cluster-specific transmission rates to those observed through analysis of NHSS data. We simulated molecular cluster detection, then determined the full extent and composition of these clusters, including undiagnosed infections and infections without a reported sequence, which cannot be observed in real-world data. Additionally, we calculated HIV transmission rates in simulated cluster networks, to assess whether high rates of transmission observed in detected clusters are observed across full cluster networks.

## Methods

### Model

We used the Progression and Transmission of HIV (PATH 4.0), a stochastic dynamic network simulation model of sexual transmission of HIV in the United States; details of this model have been previously described<sup>9</sup>. In brief, this model utilizes an agent-based evolving network modeling (ABENM) simulation technique, which includes both individual-level and compartmental modeling structures. Persons with HIV (PWH) and their immediate contacts are simulated at the individual level, and all other susceptible persons, using a compartmental modeling structure. Immediate contacts are defined as all sexual partners a person with HIV will have over their lifetime. We modeled ‘pseudo’-geographic jurisdictions to capture mixing within and across jurisdictions<sup>10</sup>, and geographic heterogeneity in diagnosis and viral suppression. The compartmental model specifies age-group, transmission risk-group, number of lifetime partners, and pseudo-geographic jurisdictions. The individual model specifies age, transmission risk-group, lifetime partners, pseudo-geographic jurisdiction, and HIV infection and care status.

Details of model implementation including modules governing network formation, transmission, and disease progression are described in the supplemental materials. Briefly, the model was initialized with a population representative of PWH infected through sexual transmission in the United States as of 2006. We ran the simulation from 2006 to 2017 in monthly-time steps, accounting for the changes in testing and care behaviors in each year being simulated. Molecular cluster detection was replicated as of model year 2017 to identify clusters among persons with HIV diagnosed during 2015–2017. In the model, we defined genetic distance (GD) between any two HIV infections as the sum of the difference between each infection’s time of diagnosis and time of the transmission event that commonly connects them in a transmission network (Supplemental Figure 1). Then, for each GD threshold, we generated clusters by grouping infections such that each infection in a cluster is within the specified GD threshold of at least one other infection in the cluster. We set GD thresholds by assuming that the viral sequence will change approximately 1%, on average, over a 10-year interval<sup>11</sup>, e.g., a GD threshold of 5 years in the simulation model would represent a molecular genetic distance of 0.5%. Because HIV changes independently in each person following a transmission event, a 0.5% threshold corresponds to transmission events that occurred ~2–3 years prior.

### Transmission rates

We calculated the transmission rate of a cluster as the number of new infections in the cluster during 2015–2017 divided by the cumulative person-time during 2015–2017 in which persons in the cluster were infected with HIV (see supplemental materials for additional detail).

### Expanded validation of transmission network, cluster distributions, and transmission rate

We compared selected model clustering outcomes to cluster outcomes for persons with HIV diagnosed during January 2015–December 2017 for whom a nucleotide sequence data was reported to NHSS by December 2017, excluding persons for whom injection drug use was identified as the transmission category. We compared the proportion of diagnosed infections that were clustered, and the distribution of cluster sizes across genetic distance thresholds of 0.5%, 0.75%, 1.0%, 1.5%, and 2.0% (equivalent to 5, 7.5, 10, 15, and 20 years of distance in the model, respectively). Additionally, transmission rates in detected clusters were compared between model simulations and as observed in NHSS data. Results of this expanded validation are reported in the supplemental materials.

### Cluster composition metrics

In the model, we assessed the full extent and composition of clusters defined at the 0.5% genetic distance threshold, CDC's threshold for routine HIV cluster detection<sup>5</sup>. To do so, we first identified 'detected clusters' in the model, which reflect the portion of a full cluster that can be identified through analysis of NHSS data, based on diagnosed infections for which a sequence is reported (see supplemental materials for more detail).

### Detected and full clusters

To replicate the incomplete reporting of HIV sequence data for diagnosed infections, we randomly selected 48% of simulated diagnosed infections and set their status as a 'sequenced' infection (approximating the completeness of sequence reporting in NHSS during 2015–2017<sup>5</sup>). To identify detected clusters, the cluster generation algorithm is applied to only these sequenced infections that were diagnosed during 2015–2017.

To understand the full extent and composition of clusters in the model, for each detected cluster, we identified the full cluster that would have been observed if all HIV infections in the cluster were diagnosed and sequenced (Supplemental figure 1). To generate full clusters in the model we applied the cluster generation algorithm to all infections diagnosed during 2015–2017 as well as all infections that were undiagnosed as of the last model time step, assuming that sequences were available for all infections. Infections diagnosed prior to 2015 that linked to the cluster at the 0.5% threshold were also included. Detected clusters were mapped to the corresponding full clusters to allow for comparisons of detected to full clusters.

Attributes of infections within a cluster and overall, such as recency of infection and diagnosis status, were determined as of the last model time step.

Priority clusters: To focus on clusters with active growth, we further identified ‘priority clusters’ following CDC’s national priority cluster definition, defined as detected clusters at the 0.5% genetic distance threshold with at least 3 infections diagnosed and sequenced in the most recent year of analysis<sup>5</sup>. Non-priority clusters were those clusters that did not meet this criterion.

Because routine analyses of NHSS data to identify priority clusters are conducted prospectively by health departments and CDC, priority clusters are identified when they first meet the priority cluster definition. To approximate this in the model we distinguished new and ongoing priority clusters as of the time of the last model timestep (December 2017).

Ongoing priority clusters: Defined as clusters with at least 3 diagnosed and sequenced infections in 2017 in which at least 3 additional infections in the cluster were diagnosed and sequenced in a single calendar year during 2012–2016 (indicating that the cluster first met priority criteria prior to 2017).

New priority clusters: Defined as those with at least 3 diagnosed and sequenced infections in 2017 in which fewer than 3 infections in the cluster were diagnosed and sequenced in each calendar year during 2012–2016 (indicating that the cluster met priority criteria for the first time in 2017).

### **Model replications**

We ran the simulation 30 times. Within each model run, we conducted 11 replications in which the subset of 48% of ‘sequenced’ infections were randomly selected from the simulated diagnosed infections. We calculated the median and range for cluster outcomes from the 330 sampled scenarios.

## **Results**

### **Expanded validation of transmission network, cluster distributions, and transmission rate**

Cluster distributions generated by the simulation model aligned with distributions observed through analysis of NHSS data across a range of genetic distance thresholds (Supplemental Figure 2a and 2b). Additionally, transmission rates in detected clusters simulated in the model were similar to those observed in clusters identified through analysis of NHSS data. Full expanded validation results are described in the supplemental materials.

### **Overall clustering at the 0.5% genetic distance threshold**

Across all simulations at the 0.5% genetic distance threshold, a median of 25.4% (range, 11.9–32.7%) of infections diagnosed during 2015–2017 with a sequence were in a detected cluster. Among these diagnosed and sequenced infections in detected clusters, 2.5% (0.0–13.5%) were in a new priority cluster, 5.9% (0.0–28.1%) in an ongoing priority cluster, and 90.8% (69.3–100%) in a non-priority cluster.

### Composition of detected clusters by recency of infection

Diagnosed infections detected as part of a cluster in model simulations were more often acquired recently than infections overall (Figure 1). A median of 36.5% (range 31.6–41.7%) of infections diagnosed during 2015–2017 with a sequence in detected clusters had been acquired within the last 3 years; in new priority clusters, the median was 58.2% (range 30.1–100.0%). By contrast, a median of 21.3% (range 18.7–26.9%) of infections diagnosed during the same period with a sequence not in a cluster had been acquired within the last 3 years.

### Extent of full clusters

In addition to infections detected as part of a cluster, full clusters included additional infections that were undiagnosed, infections that were diagnosed during 2015–2017 but didn't have a sequence, and infections diagnosed prior to 2015.

Across all categories and sizes at detection, full clusters were substantially larger than detected clusters, ranging from 3–9 times the detected cluster size (Table 1). While the median size of newly detected priority clusters was 4 (range 3–9), the corresponding full clusters were a median of size 14 (range 3–74). Ongoing priority clusters were the largest, with a median detected size of 11 (range 3–33), corresponding to the largest full clusters of a median size 94 (range 7–318).

### Composition of full clusters by diagnosis status and recency of infection

Compared to the infections in the model that were not clustered at the 0.5% genetic distance threshold, infections in clusters were more often undiagnosed (Figure 2a). New priority clusters had the greatest proportion of infections that were undiagnosed, with a median of 36.3% (range 11.1–72.6%) while this proportion declined in ongoing priority clusters to a median of 22.1% (range 9.6–37.0%). For each diagnosed infection detected as part of a new priority cluster, there were a median of an additional 1.4 undiagnosed infections; for ongoing priority clusters, this ratio was 1.7.

While infections diagnosed prior to 2015 made up the greatest proportion of non-clustered infections in the simulation (median 73.7%, range 72.5–74.7%), they represented a smaller proportion in all clusters, with a median of 35.7% (range 26.1–46.1%) (Figure 2a). New priority clusters had the lowest proportion of these older diagnosed infections, though with high variability, with a median of 26.4% (range 0–54.3%). Among clusters, ongoing priority clusters had the highest proportion of older diagnosed infections, with a median of 55.2 (range 34.5–79.5).

Among undiagnosed infections, those that were part of a cluster had more frequently been acquired within the last 3 years (median 79.0%, range 73.4–84.3%) compared to a median of 57.7% (range 54.2–61.2%) of undiagnosed infections not in clusters (Figure 2b).

### Transmission rates in full clusters

Transmission rates across full clusters defined at the 0.5% genetic distance threshold were in similar ranges to transmission rates observed in detected clusters across all categories

(Figure 3). Transmission rates in new priority clusters were highest compared to other categories, with a median transmission rate in the full clusters of 19.9 transmission events per 100 person-years, ranging from 7.6 to 59.5 transmission events per 100 person-years across simulations. This reflects a median transmission rate of >4 times the overall rate of 4.5 (range 4.3–5.0) transmission events per 100 person-years among all HIV transmissions within the entire simulation (including non-clustered as well as clustered infections), with medians across model runs ranging from 2–13 times the overall rate.

## Discussion

This nationally representative simulation model successfully replicates the transmission dynamics that generate large HIV clusters observed through surveillance. This provides a unique opportunity to understand the full extent and composition of clusters, allowing a systematic understanding of key questions to guide HIV CDR strategies. The model suggests that CDC's current cluster detection approaches disproportionately capture recently acquired infections, consistent with other evidence that these clusters identify the leading edge of transmission<sup>1,5</sup>. The substantially larger full size of these clusters as compared to the size as detected through surveillance data (from 3 to 9 times the detected cluster size), along with the disproportionate proportion of undiagnosed infections (36.3%) and the high transmission rates in these clusters (19/100 person-years) indicate opportunities for public health intervention and impact.

The substantially larger full extent of clusters observed in the model is consistent with observations from selected field investigations, in which response activities have identified previously diagnosed infections that are linked to clusters and testing in cluster networks has uncovered previously undiagnosed infections<sup>1,7</sup>. These findings reinforce that cluster interventions focused solely on individuals detected as part of a cluster, such as prioritizing detected cluster members for linkage-to-care activities, are insufficient. Effective strategies to reach and understand the full network impacted by rapid transmission are crucial. While partner services is an important prevention tool to reach networks, it has important limitations, including limitations in clients' ability or willingness to name partners<sup>12</sup>. Response strategies to reach and understand the full network, such as social and sexual network testing strategies, rapid qualitative assessments, and enhanced partner services, are important to guide interventions at the individual, network, and structural levels<sup>1</sup>.

The model further reveals the composition of full clusters, allowing us to better understand the networks underlying clusters, assess the potential for public health impact, and guide intervention strategies. Detected clusters defined using the 0.5% genetic distance threshold were disproportionately associated with infections acquired more recently. This is consistent with what has been observed through analysis of indicators of recency available in surveillance data<sup>5</sup>, but the model allows us to observe this systematically, while indicators of recency in surveillance data are frequently unavailable.

Undiagnosed infections were disproportionately associated with clusters, with notably high proportions of infections in new priority clusters being undiagnosed and >1.4 additional undiagnosed infections for each detected infection. These undiagnosed infections in clusters

were also more likely to have been acquired recently (i.e., in the prior 3 years). This indicates that focused efforts to ensure HIV testing resources reach these networks may have a disproportionate impact on increasing awareness of infection and receipt of HIV care. Additionally, because undiagnosed infections contribute disproportionately to ongoing transmission<sup>13</sup>, increasing testing in cluster networks may accelerate the prevention of new HIV infections. Focused HIV testing in networks of ongoing priority clusters may also have a disproportionate impact on increasing awareness, with >1.7 additional undiagnosed infections for each detected infection in these clusters.

The full extent of clusters is smaller when detected clusters first meet priority criteria, and substantially larger for ongoing priority clusters. This suggests that early intervention, when clusters first meet priority criteria, can more efficiently reach networks, and potentially prevent future infections. Particularly high transmission rates when clusters are first detected further underscore the importance of this early intervention. High transmission rates have been observed in newly detected clusters through NHSS<sup>2,5</sup>, and our findings reveal that those elevated transmission rates among detected clusters are not artifacts of detection. Transmission rates in full clusters, that included those infections that are undiagnosed, those diagnosed without a sequence available, and those diagnosed prior to 2015–2017, were similarly high.

A nationally representative simulation model is important to effectively assess HIV CDR in the United States. Transmission of HIV does not respect jurisdictional boundaries<sup>10</sup>, and rapidly growing clusters in the U.S. frequently involve multiple states<sup>2</sup>. Moreover, the inclusion of cross-state transmission in our model was found to be crucial to replicating the dynamics that generate large clusters observed through U.S. NHSS data<sup>9</sup>. The novel simulation technique implemented here, which combines agent-based and compartmental techniques, was critical to allow a nationally representative simulation to be computationally feasible. With the relatively low HIV prevalence in the United States, the computational challenges in traditional agent-based network models (in which the full populations are simulated in the individual-level simulation) make them insufficient to generate large clusters. Additionally, the robust validation of clustering proportions and cluster size distributions across multiple genetic distance thresholds, in addition to the validation of transmission rates in clusters, suggests that the model is successfully replicating the dynamics, including network structure, sexual behaviors, and care behaviors, that give rise to rapidly growing clusters that have resulted in large scale public health response activities<sup>1</sup>.

This model analysis has limitations. First, in these model runs clusters were identified at a single time point, the last time step of the model. In practice, at health departments and CDC, cluster detection is conducted routinely at monthly or quarterly intervals; once identified, priority clusters are tracked over time. Therefore, our finding that ongoing priority clusters had a lower proportion of infections detected (i.e., diagnosed in the prior 3 years and with a sequence) than other cluster categories is not surprising; in public health practice, those previously diagnosed infections would have been detected as part of the cluster previously, when they first met the criteria for ‘new priority clusters’. Second, we focused on 2015–2017, given the availability of key model input parameters for calibration of the care continuum. Although sequence completeness in NHSS during this time period

was lower nationwide due to participation of only 27 jurisdictions representing 70% of HIV diagnoses in 2015, we assumed that the clustering dynamics in these 27 jurisdictions were representative of all jurisdictions, and our data suggest that this assumption is a reasonable one (unpublished data). Finally, this model simulates only sexual transmission of HIV, which is the predominant route of transmission for rapidly growing clusters in the United States<sup>2,6</sup>, and therefore not a major limitation. Future inclusion of other modes of transmission would allow the model to address questions related to public health response to clusters attributable to injection drug use.

This model provides a robust foundation to assess key questions related to HIV CDR, and our findings reinforce the importance of detecting and responding to clusters early. Validation of model outcomes against real-world observations gives us confidence that the model replicates the transmission dynamics that generate rapidly growing clusters. These findings demonstrate the potential for response to rapid transmission clusters identified through CDC's molecular cluster detection methods to have a disproportionate contribution to the diagnosis of previously unrecognized HIV infections and prevention of ongoing HIV transmission.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

We would like to acknowledge the contributions of Katherine Hicks, Justin Carrico, and Claire Mellott (RTI Health Solutions) to this work under Project No. 0216054.004, CDC Contract No. 200-2017-93666.

## Conflicts of Interest and Source of Funding:

CG was supported by a grant from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI127236. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

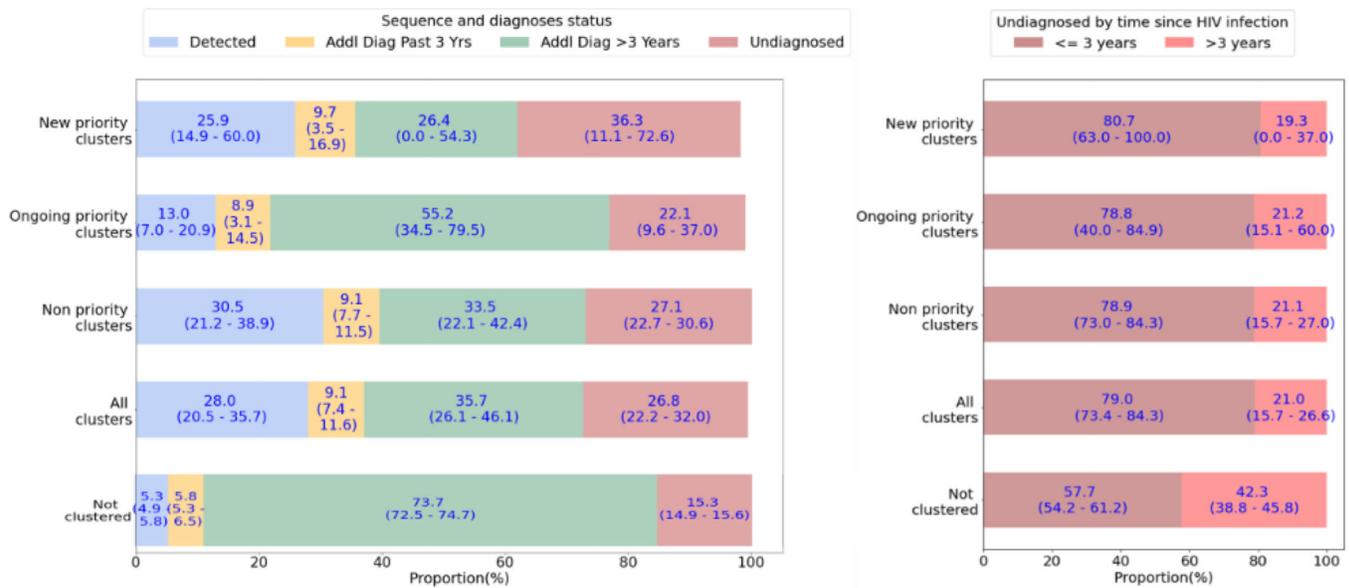
## References

- Oster AM, Lyss SB, McClung RP, et al. HIV cluster and outbreak detection and response: the science and experience. *Am J Prev Med.* Nov 2021;61(5 Suppl 1):S130–S142. doi:10.1016/j.amepre.2021.05.029 [PubMed: 34686282]
- Perez SM, Panneer N, France AM, et al. Clusters of rapid HIV transmission among gay, bisexual, and other men who have sex with men - United States, 2018–2021. *MMWR Morb Mortal Wkly Rep.* Sep 23 2022;71(38):1201–1206. doi:10.15585/mmwr.mm7138a1 [PubMed: 36136909]
- 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.* Dec 1 2015;70(4):444–51. doi:10.1097/QAI.0000000000000809 [PubMed: 26302431]
- Kosakovsky Pond SL, Weaver S, Leigh Brown AJ, Wertheim JO. HIV-TRACE (TRANsmission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens. *Mol Biol Evol.* Jul 1 2018;35(7):1812–1819. doi:10.1093/molbev/msy016 [PubMed: 29401317]
- Oster AM, France AM, Panneer N, et al. Identifying clusters of recent and rapid HIV Transmission through analysis of molecular surveillance data. *J Acquir Immune Defic Syndr.* Dec 15 2018;79(5):543–550. doi:10.1097/QAI.0000000000001856 [PubMed: 30222659]

6. Oster AM, Panneer N, Lyss SB, et al. Increasing capacity to detect clusters of rapid HIV transmission in varied populations-United States. *Viruses*. Mar 30 2021;13(4)doi:10.3390/v13040577
7. Analise C, Monterosso SM, Samuel Goings, Allyson Morris, Anne Marie France, Sharoda Dasgupta, Alexandra Oster, Miranda Fanning. Identifying and Investigating a Rapidly Growing HIV Transmission Cluster in Texas. presented at: Conference on Retroviruses and Opportunistic Infections; February 13, 2017 2017; Seattle, WA. <https://www.croiconference.org/abstract/identifying-and-investigating-rapidly-growing-hiv-transmission-cluster-texas/>
8. Murray M. Determinants of cluster distribution in the molecular epidemiology of tuberculosis. *Proc Natl Acad Sci U S A*. Feb 5 2002;99(3):1538–43. doi:10.1073/pnas.022618299 [PubMed: 11818527]
9. Singh S, France AM, Chen YH, Farnham PG, Oster AM, Gopalappa C. Progression and transmission of HIV (PATH 4.0)-A new agent-based evolving network simulation for modeling HIV transmission clusters. *Math Biosci Eng*. Mar 3 2021;18(3):2150–2181. doi:10.3934/mbe.2021109 [PubMed: 33892539]
10. Board AR, Oster AM, Song R, et al. Geographic Distribution of HIV Transmission Networks in the United States. *J Acquir Immune Defic Syndr*. Nov 1 2020;85(3):e32–e40. doi:10.1097/QAI.0000000000002448 [PubMed: 32740373]
11. Hightower GK, May SJ, Perez-Santiago J, et al. HIV-1 clade B pol evolution following primary infection. *PLoS One*. 2013;8(6):e68188. doi:10.1371/journal.pone.0068188
12. Magaziner S, Montgomery MC, Bertrand T, et al. Public health opportunities and challenges in the provision of partner notification services: the New England experience. *BMC Health Serv Res*. Jan 31 2018;18(1):75. doi:10.1186/s12913-018-2890-7 [PubMed: 29386023]
13. Li Z, Purcell DW, Sansom SL, Hayes D, Hall HI. Vital Signs: HIV transmission along the continuum of care - United States, 2016. *MMWR Morb Mortal Wkly Rep*. Mar 22 2019;68(11):267–272. doi:10.15585/mmwr.mm6811e1 [PubMed: 30897075]

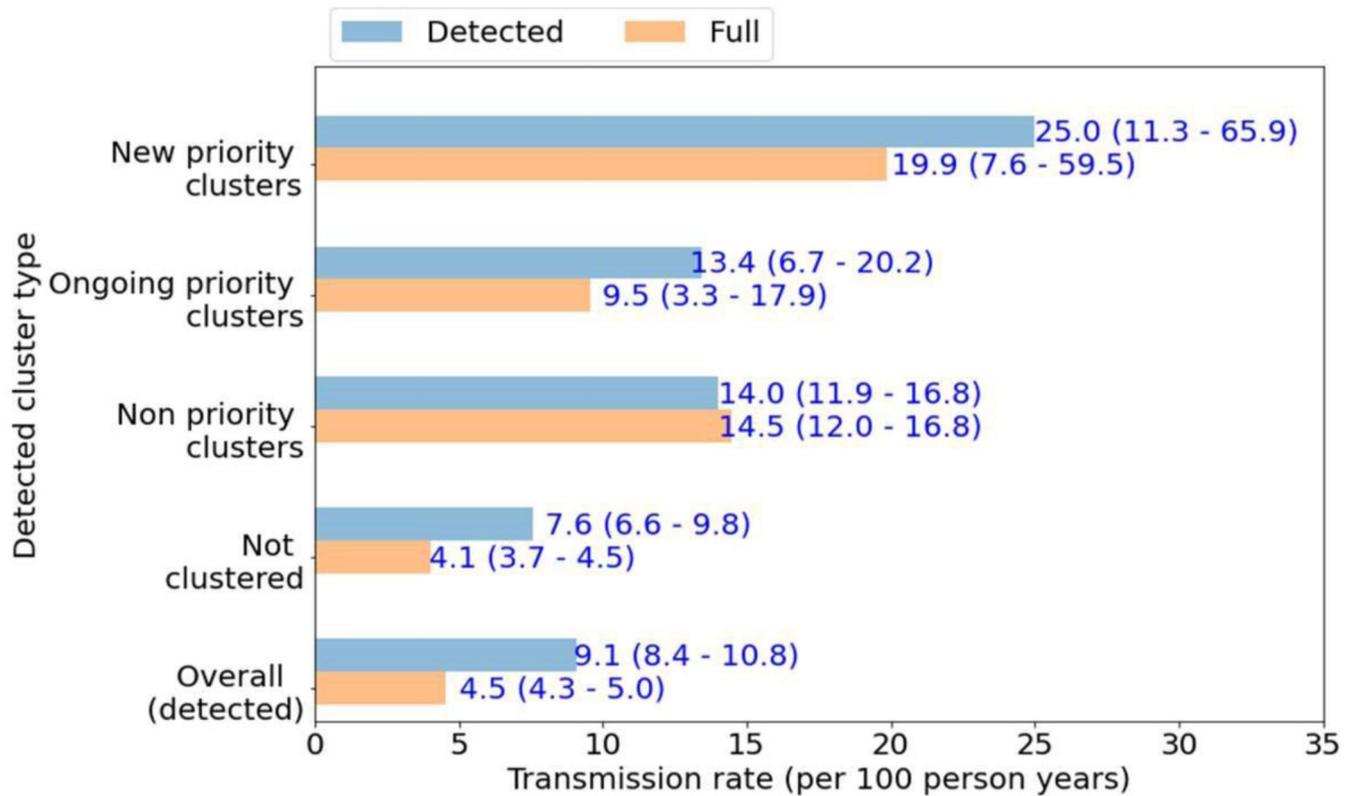


**Figure 1.** Simulated proportions of detected (diagnosed and sequenced) HIV infections by time since HIV infection and cluster status at time of cluster detection (12/31/2017), among simulated persons with HIV infection diagnosed during 2015–2017, by cluster status defined at 0.5% genetic distance threshold.



**FIGURE 2:**

A) Median proportion and range of infections in full clusters by diagnosis and sequencing status, among clusters defined at 0.5% genetic distance threshold; B) Median proportion and range of undiagnosed infections in full clusters by time since HIV infection as of the last time step of the model simulation (December 2017)



**FIGURE 3:** Median and range of transmission rates in molecular detected and corresponding full clusters by cluster type, as generated in the simulation model with clusters defined at the 0.5% genetic distance threshold.

**Table 1.**

Size of detected clusters and full clusters (median, min, and max) across model simulations, by cluster type as generated in the simulation model with clusters defined at the 0.5% genetic distance threshold.

Cluster type	Detected cluster* size (Median, min, max)	**Full cluster size (Median, min, max)
New priority clusters	4 (3–9)	14 (3–74)
Ongoing priority clusters	11 (3–33)	94 (7–318)
Non-priority clusters	2 (2–19)	5 (2–183)
All clusters	2 (2–33)	5 (2–318)

\* Detected clusters represent the portion of a cluster that can be identified through analysis of reported surveillance data (i.e., infections that were diagnosed during 2015–2017 with a sequence reported), assuming 48% sequence completeness.

\*\* In addition to infections detected as part of a cluster, full clusters include additional infections that were undiagnosed, infections that were diagnosed during 2015–2017 but didn't have a sequence, and infections diagnosed prior to 2015.