

RESEARCH ARTICLE

Uncovering the impact of randomness in HIV hotspot formation: A mathematical modeling study

Nao Yamamoto¹, Daniel T. Citron¹, Samuel M. Mwalili², Duncan K. Gathungu^{2,3}, Diego F. Cuadros⁴, Anna Bershteyn^{1*}

1 Department of Population Health, New York University Grossman School of Medicine, New York, New York, United States of America, **2** Strathmore University, Ole Sangale Road, Nairobi, Kenya, **3** Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, **4** Digital Epidemiology Laboratory, Digital Futures, University of Cincinnati, Cincinnati, Ohio, United States of America

* Anna.Bershteyn@nyulangone.org



Abstract

Background

HIV hotspots, regions with higher prevalence than surrounding areas, are observed across Africa, yet their formation and persistence mechanisms remain poorly understood. We hypothesized that random fluctuations during the early stages of the HIV epidemic (1978–1982), amplified by positive feedback between HIV incidence and prevalence, play a critical role in hotspot formation and persistence. To explore this, we applied a network-based HIV transmission model, focusing on randomness in the spatial structure of the epidemic.

Methods

We adapted a previously validated agent-based network HIV transmission model, EMOD-HIV, to simulate HIV spread in western Kenya communities. The model includes demographics, age-structured social networks, and HIV transmission, prevention, and treatment. We simulated 250 identical communities, introducing stochastic fluctuations in network structure and case importation. Outliers were defined as communities with prevalence > 1.5x the median, and persistence as meeting these criteria for >70% of 1980–2050. We systematically varied community size (1,000–10,000), importation timing (1978–1982), and importation patterns (spread over 1, 3, or 5 years), and calculated the proportion of outliers and persistent outliers.

Results

HIV prevalence outliers were more common in smaller communities: in 1990, 25.3% (uncertainty interval: 22.3%–28.2%) of 1,000-person communities vs. 9.1% (uncertainty interval: 6.9%–11.4%) of 10,000-person communities. By 2050, 21.6%

OPEN ACCESS

Citation: Yamamoto N, Citron DT, Mwalili SM, Gathungu DK, Cuadros DF, Bershteyn A (2025) Uncovering the impact of randomness in HIV hotspot formation: A mathematical modeling study. *PLoS Comput Biol* 21(6): e1013178. <https://doi.org/10.1371/journal.pcbi.1013178>

Editor: Nicholas Geard, University of Melbourne, AUSTRALIA

Received: November 2, 2024

Accepted: May 29, 2025

Published: June 16, 2025

Copyright: © 2025 Yamamoto et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data for this study were generated through the EMOD-HIV simulation model and are available within the manuscript and its Supporting Information files. The EMOD software, along with documentation and details on how to generate similar simulated data, can be accessed

at <https://docs.idmod.org/projects/emodpy-hiv/en/latest/index.html>. This study used simulated data, so no human subjects or personal data were involved.

Funding: Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R01MH130238 and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Numbers R01AI174932 and R01AI179417 awarded to A.B. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: AB discloses grants to her institution from the US National Institutes of Health, New York City Department of Health and Mental Hygiene, Bill and Melinda Gates Foundation, and Foundation for Innovative New Diagnostics, and consulting fees from Gates Ventures. All other authors declare that they have no competing interests.

of 1,000-person communities were persistent outliers, compared to none in larger communities. Autocorrelation of HIV prevalence was high (Pearson's correlation coefficient 0.801 [95% CI: 0.796–0.806] for 1,000-person communities), reflecting feedback that amplified early fluctuations.

Conclusions

Early random fluctuations contribute to the emergence and persistence of prevalence outliers, especially in smaller communities. Recognizing the role of randomness in prevalence outlier formation in these settings is crucial for refining HIV control strategies, as traditional methods may overlook these areas. Adaptive surveillance systems can enhance detection and intervention efforts for HIV and future pandemics.

Author summary

In this study, we explored how random events early in the HIV epidemic may have contributed to the formation of HIV prevalence outliers. Using a simulation model of HIV spread in communities in western Kenya, we introduced random variations in the initial spread of the disease and examined how these fluctuations affected infection patterns over time. We found that community-level prevalence exceeding 1.5 times the median could emerge through random variations in an appreciable fraction of small communities, and that outliers were more likely to persist over time in such settings. In contrast, larger communities appeared to need additional factors beyond random chance for outliers to form. Our findings suggest that random fluctuations can influence the development of HIV prevalence outliers, especially in smaller communities. Recognizing this could enhance public health strategies by encouraging more detailed monitoring of HIV in vulnerable areas. By understanding the role of randomness, we can refine approaches to detecting and controlling the spread of HIV and other infectious diseases in the future.

Introduction

Four decades into the HIV/AIDS pandemic, there remains an incomplete understanding of why certain communities are disproportionately affected by the virus [1–2]. Despite extensive research, two-thirds of people living with HIV (PLHIV) reside in Africa [3], with the eastern and southern regions bearing the brunt of the epidemic. At the national level, the Kingdom of Eswatini has the world's highest HIV prevalence, affecting 27% of adults [4]. Sub-nationally, similar high prevalence rates are found along the Lake Victoria coast in Eastern Africa [5], particularly within fishing communities [6–8], and South Africa's KwaZulu-Natal Province [9]. At finer, community-level scales, HIV prevalence [10–13] and incidence [14] exhibit significant variation, sometimes differing by several-fold across neighborhoods just a few kilometers apart. In

KwaZulu-Natal, door-to-door surveys have revealed that HIV prevalence can differ more than three-fold between isigodis (neighborhoods) located just 1–10 kilometers apart [15–17]. While individuals often select sexual partners within their own isigodi [10,18], reinforcing localized transmission dynamics, evidence also suggests that migration plays a role in shaping these patterns, as PLHIV may be more likely to relocate to high-prevalence communities [19].

These stark differences in HIV prevalence, observable across various spatial scales, resemble fractal patterns—mathematical structures that exhibit self-similarity, meaning similar patterns recur at different levels of magnification [20–23]. Fractals frequently occur in nature [24,25], suggesting that further study of HIV’s spatial variation could yield new insights into the epidemic’s dynamics. A range of behavioral and biological mechanisms have been proposed to explain this variability [26,27], including the formation of “hotspots”—areas with anomalously high HIV prevalence compared to neighboring regions [28]. In some cases, hotspots are linked to behaviors and situations that increase HIV risk, such as transactional sex [29,11], cultural practices like polygamy [30], or occupations involving extended separation from partners, such as mining [31,32], long-distance trucking [33], or agricultural work [34]. Biologically, hotspots often coincide with areas where male circumcision rates are lower [35,36], although recent public health initiatives have increased circumcision coverage in many communities [36]. Additionally, hotspots frequently overlap with areas of high prevalence of other sexually transmitted infections (STIs) and reproductive tract infections (e.g., vaginal dysbiosis) [37]. However, the causal associations between these infections and HIV are often confounded by shared behavioral and biological risk factors.

Despite advances in understanding the drivers of the HIV epidemic, many community-level hotspots remain unexplained. In KwaZulu-Natal, the observed three-fold differences in HIV prevalence did not align with measurable differences in behavioral or biological risk factors [15–17]. Proximity to major roads was associated with some hotspots [16], suggesting that increased mobility, such as the movement of migrant workers and truck drivers, could play a role in the spatial distribution of HIV. In rural Kenya’s Nyanza counties, cultural factors, such as stigma against individuals perceived to have HIV/AIDS, often attributed to beliefs in witchcraft or family curses, may also contribute to the classification of certain geographical areas as hotspots [38–41].

The phenomenon of “patchy” and seemingly random epidemic hotspots is not unique to HIV; it has been observed across various infectious diseases and is often associated with stochastic processes in finite populations. Such fluctuations tend to be more pronounced in smaller populations [42,43]. Given HIV’s chronic nature, we hypothesize that stochastic fluctuations in early HIV case introductions may have led to the sustained patchy patterns observed today [44]. Despite this broader phenomenon, there is no universally accepted threshold for defining HIV “hotspots.” Some definitions use absolute prevalence cutoffs, while others compare prevalence within a region to its surroundings. For instance, hotspots have been defined as communities where HIV prevalence is at least twice that of neighboring areas, as observed in fishing communities around Lake Victoria [28], while others report even greater disparities, such as a 2.5-fold difference in prevalence between high-burden fishing communities and surrounding regions [45]. However, a strict threshold may not always capture the full spectrum of spatial heterogeneity, particularly in settings with fine-scale prevalence variations. Given this variability, we define HIV prevalence outliers based on a relative threshold (1.5x the median prevalence across simulated communities) and conduct sensitivity analyses to examine how different threshold choices influence our findings.

This study aims to assess the extent to which randomness in HIV case introductions and spread through sexual networks contributes to the formation of high-prevalence communities. We employed a stochastic, network-based HIV transmission model to simulate identical communities, varying only in random perturbations to sexual networks and initial case introductions while holding average behavioral and biological risk parameters constant. By systematically varying population size, intensity, and timing of case introduction, we examined conditions under the formation of communities with high HIV prevalence could arise under random fluctuations in initial seeding events. Through this approach, we establish a baseline for understanding the degree of variability that can arise due to chance alone. This baseline can be useful in assessing which deviations in prevalence are unusual for a community’s size and intensity of case importation,

and therefore potentially indicative of epidemiological risk factors for HIV transmission. We further assessed the sensitivity of the observed patterns to alternative methods of outlier detection and thresholds that define an outlier. This work represents, to our knowledge, the first study exploring when and to what extent HIV prevalence outliers can arise purely by chance, independent of underlying behavioral or biological risk factors. By recognizing the role of stochastic processes, public health strategies can better account for the possibility of randomly emerging high-prevalence areas.

Results

Overall prevalence

Across the 93,750 simulations used in the analysis, mean HIV prevalence was 20.1%, with a standard deviation (SD) of 5.8%, in 2018, which was within the uncertainty range of HIV prevalence in Kisumu County estimated in the same year via population-based surveillance (17.5%, 95% CI: 13.6–21.3%). Mean HIV prevalence differed by initial community (i.e., social network) size (S1 Table), with prevalence in 2018 at 17.2% (uncertainty interval: 17.0–17.3%) for communities with an initial population of 1,000 and 21.4% (uncertainty interval: 21.3–21.4%) for communities with an initial population of 10,000.

Population size

Communities with smaller initial population sizes exhibited systematically larger variance in HIV prevalence across all timepoints. Fig 1 presents violin plots showing the distribution of simulated HIV prevalence in 2018 for communities with varying initial population sizes (1,000; 3,000; 5,000; 7,000; and 10,000 individuals). These plots highlight the variability in prevalence across simulations and compare them to the observed empirical HIV prevalence in Kisumu County for the same year. Notably, the spread of HIV prevalence is broader in smaller populations, especially in communities with an initial population size of 1,000. This increased variability underscores how random fluctuations in transmission dynamics can lead to substantial differences in prevalence within smaller populations.

All simulated epidemics peaked in the mid-2000's and declined by 2050, reflecting HIV treatment and prevention scale-up. Although absolute variance in HIV prevalence was highest when epidemics peaked (Fig 2), the relative variance (i.e., variability in prevalence relative to the mean prevalence) was highest during the early HIV epidemic (e.g., 1990) and as the HIV epidemic became increasingly controlled (e.g., 2050). In panels (d-f) in Fig 2, each scatter plot represents the relative HIV prevalence at various time points, with the horizontal line marking the threshold of 1.5 times the median prevalence used to define HIV prevalence outliers. A higher proportion of simulations for smaller populations cross this threshold, particularly in the early years of the epidemic. This observation suggests that randomness in early transmission events, in the absence of additional importations, results in greater divergence in prevalence estimates across simulations, especially in smaller communities where random variations have a more significant impact. As population size increases, fewer simulations exceed the threshold for prevalence outliers, indicating that the effects of stochasticity are diminished in larger populations. Overall, Fig 2 demonstrates that random fluctuations in case introduction and transmission dynamics contribute to the emergence of prevalence outliers, particularly in smaller populations. In larger populations ($\geq 5,000$), the proportion of simulations exceeding the 1.5x threshold was low (<5%), suggesting a reduced influence of stochasticity alone.

Accordingly, the proportion of simulations classified as HIV prevalence outliers was highest in the early and late epidemics and lowest during the epidemic peak (Fig 3). The figure reveals that smaller communities (e.g., 1,000 individuals) consistently show a higher proportion of prevalence outlier formation across all introduction patterns, particularly in the early (1990) and late (2050) stages of the epidemic. This indicates that random fluctuations in initial seeding contribute to spatial heterogeneity in HIV prevalence, particularly in smaller communities where stochastic effects play a more pronounced role. Interestingly, during the epidemic peak (2005), the proportion of identified prevalence outliers decreased across most population sizes, suggesting that randomness plays a smaller role when HIV prevalence is high and relatively

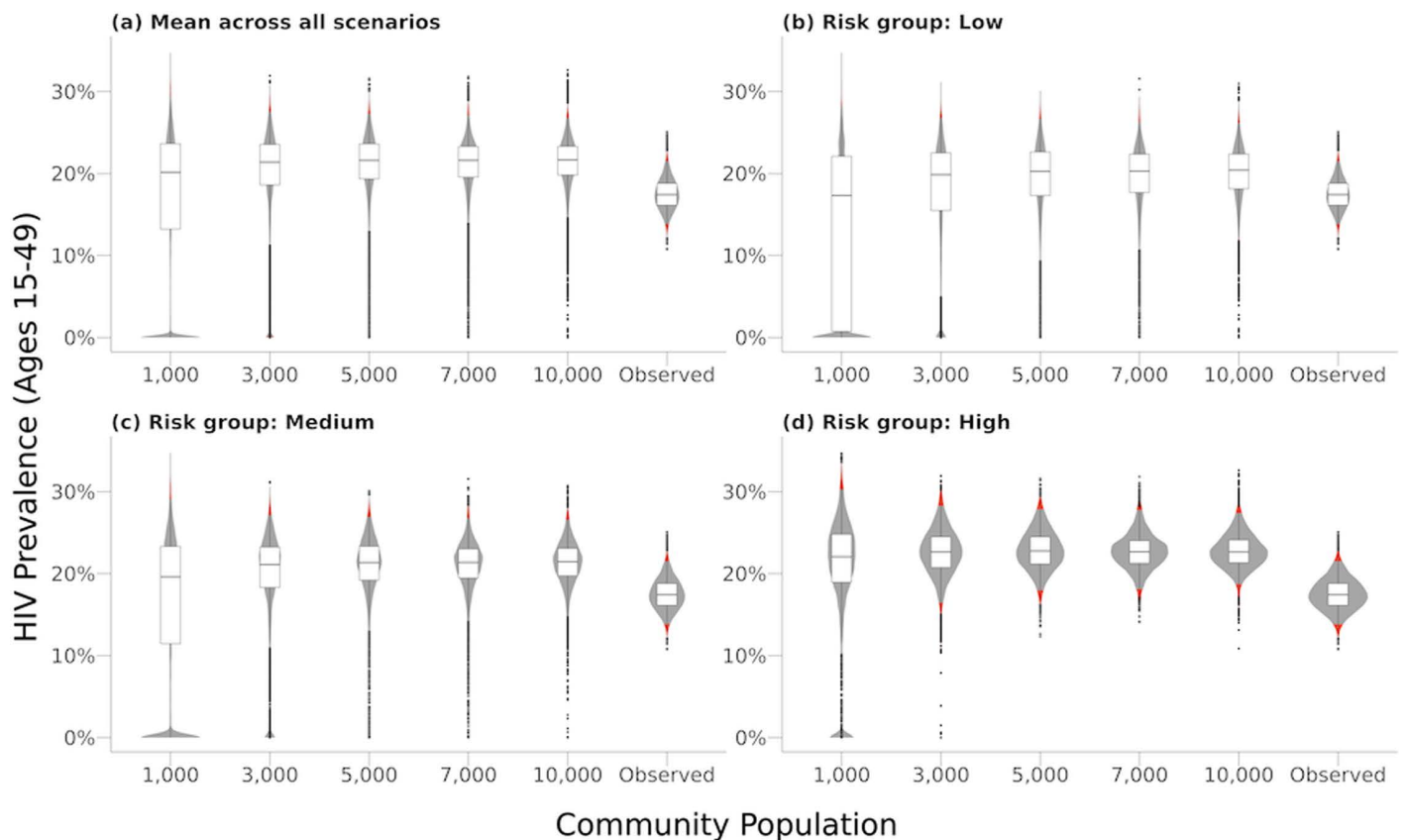


Fig 1. Violin plots of HIV prevalence in 2018 across different initial population sizes compared to the empirical HIV prevalence in Kisumu County. The distribution of simulated HIV prevalence in 2018 for various initial population sizes (1,000, 3,000, 5,000, 7,000, and 10,000 individuals) compared to empirical HIV prevalence (“Observed”) in Kisumu County in 2018. The violin plots show the distribution across 250 simulations for each community size, with boxplots indicating the median and interquartile range (IQR). Panel (a) shows all risk groups combined, and (b) presents results for low-risk groups, (c) for medium-risk groups, and (d) for high-risk groups.

<https://doi.org/10.1371/journal.pcbi.1013178.g001>

homogeneous across communities due to widespread transmission and more consistent control measures. Among the smallest communities simulated (initial population size of 1,000), which were the most prone to forming outliers, a single-year introduction of HIV cases led to 31.6% of simulations being classified as outliers in 1990, 2.0% in 2005, 2.4% in 2020, and 17.6% in 2050. For the largest communities simulated (initial population size of 10,000), a single-year introduction of HIV cases led to 0.8% of simulations being classified as outliers in 1990, no outliers observed in 2005 or 2020, and 1.6% in 2050.

The initial risk distribution of seeded individuals, as defined by the risk-stratified case importation fraction, influenced the occurrence of HIV prevalence outliers (S1 Fig). Communities where initial HIV cases occurred in higher-risk groups were more prone to becoming HIV prevalence outliers. This effect was consistent across different HIV importation patterns and initial population sizes.

Temporal patterns of outliers

The persistence of HIV prevalence outliers decreased with increasing population size (Fig 4). The columns represent the years of the first HIV introduction (ranging from 1978 to 1982), while the rows indicate various initial population sizes.

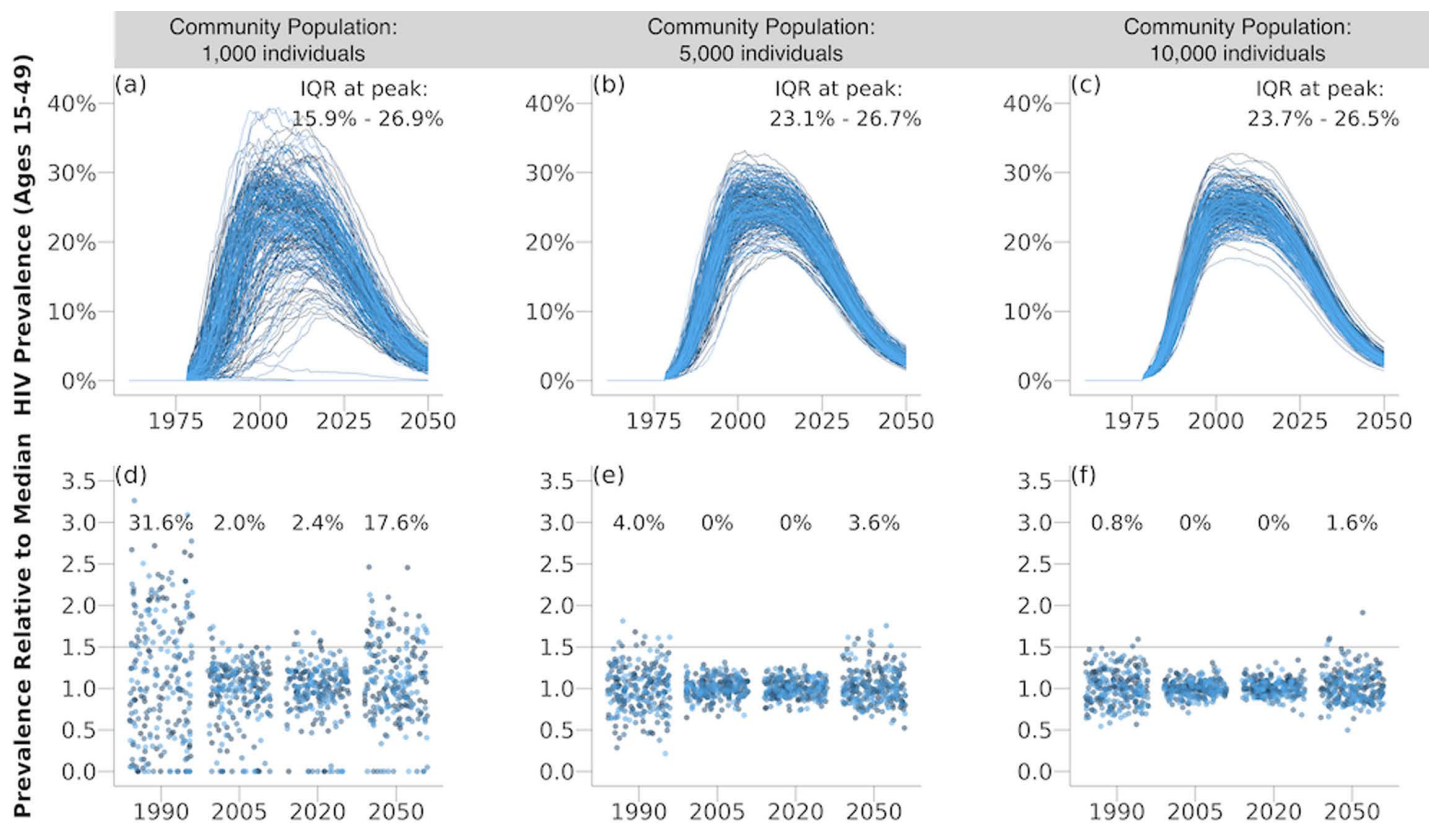


Fig 2. Impact of random fluctuations in HIV case introduction and spread on the formation of HIV prevalence outliers across different population sizes. (a-c) Ensembles of 250 simulation trajectories for population sizes of 1,000, 5,000, and 10,000 individuals, respectively. These curves illustrate the variability in epidemic trajectories due to random perturbations. The interquartile range (IQR) of peak HIV prevalence is noted for each population size. (d-f) Scatter plots showing relative HIV prevalence over time for each population size (1,000, 5,000, and 10,000 individuals). Each point represents a simulation at a specific time point, with the horizontal line indicating a relative prevalence threshold of 1.5, used to define a prevalence outlier as having a prevalence >1.5 times the median prevalence. The percentages above the scatter plots indicate the proportion of simulations that resulted in prevalence outliers for each year.

<https://doi.org/10.1371/journal.pcbi.1013178.g002>

Each cell in the heatmap displays the proportion of persistent prevalence outliers for a specific combination of introduction pattern, population size, and seeding year.

The results indicate that smaller populations (e.g., 1,000 individuals) are more likely to develop persistent prevalence outliers, particularly when the introduction of HIV cases occurs early (e.g., 1978) and is concentrated within a brief seeding period (e.g., single-year pattern). This trend is less pronounced in larger populations, suggesting that the impact of random fluctuations decreases as population size increases. Across the 375 scenarios modeled, the combination of a low demographic risk profile, earlier epidemic seeding (1978), small population (1,000), and brief seeding (single-year introduction pattern) was associated with the largest proportion of persistent HIV prevalence outliers, at 21.6%. While higher-risk groups were more likely to experience early outlier formation, low-risk populations exhibited greater outlier persistence due to lower transmission rates, which limited epidemic saturation and allowed stochastic effects to have a prolonged influence.

Autocorrelation analysis was consistent with persistence analysis in terms of high prevalence and factors associated with outlier status (Table 1). The highest autocorrelation value for outlier persistence was comparing the years 1990–2005 in a population of 1,000, with an autocorrelation of 42.4%. High autocorrelation was also observed for the pairs of years

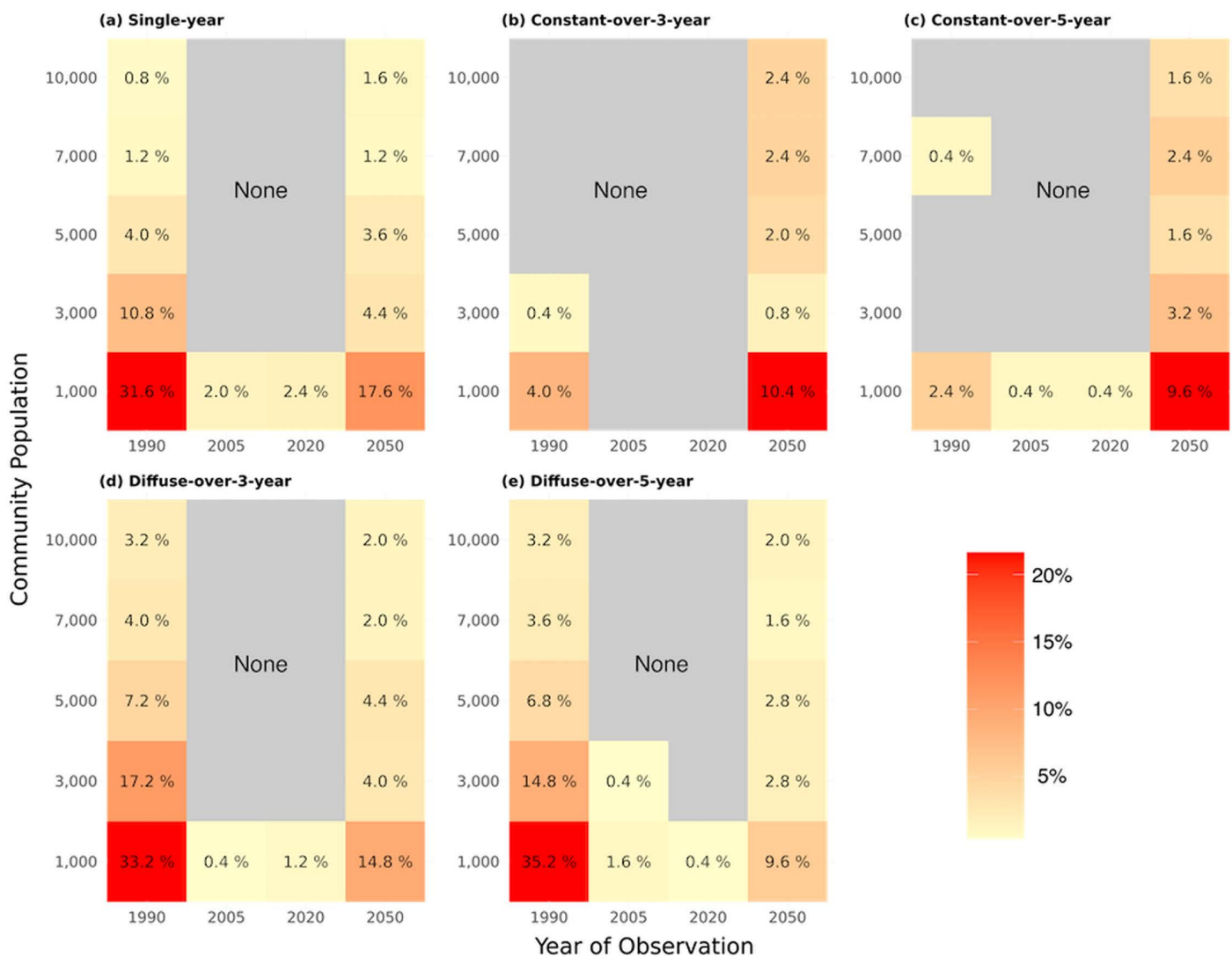


Fig 3. Proportion of communities identified as prevalence outliers across observation years, community population and HIV case introduction pattern. The proportion of communities identified as prevalence outliers for different HIV case introduction patterns and initial population sizes. The HIV case introduction patterns are: a: single-year (1 introduction event with p), b: constant-over-3-year (3 introduction events with p), c: constant-over-5-year (5 introduction events with p), d: diffuse-over-3-year (3 introduction events with $\frac{p}{3}$), and e: diffuse-over-5-year (5 introduction events with $\frac{p}{5}$). The columns represent the observation years (1990, 2005, 2020, 2050). The rows correspond to the initial population sizes. Each cell in the heatmap indicates the proportion of communities identified as prevalence outliers for a specific combination of HIV case introduction pattern, population size, and observation year.

<https://doi.org/10.1371/journal.pcbi.1013178.g003>

(2005, 2020) and (2005, 2050) in this population size, both at 23.0%. These findings underscore how smaller communities are prone to stochastic fluctuation, which can lead to the formation of persistent prevalence outliers.

Sensitivity to alternative methods of outlier detection

To further validate the persistence of outliers, we applied Density-Based Spatial Clustering of Applications with Noise (DBSCAN) as an alternative method to detect outliers. DBSCAN identifies outliers based on local data point density, allowing for the detection of simulations with prevalence rates that deviate from the majority, independent of any

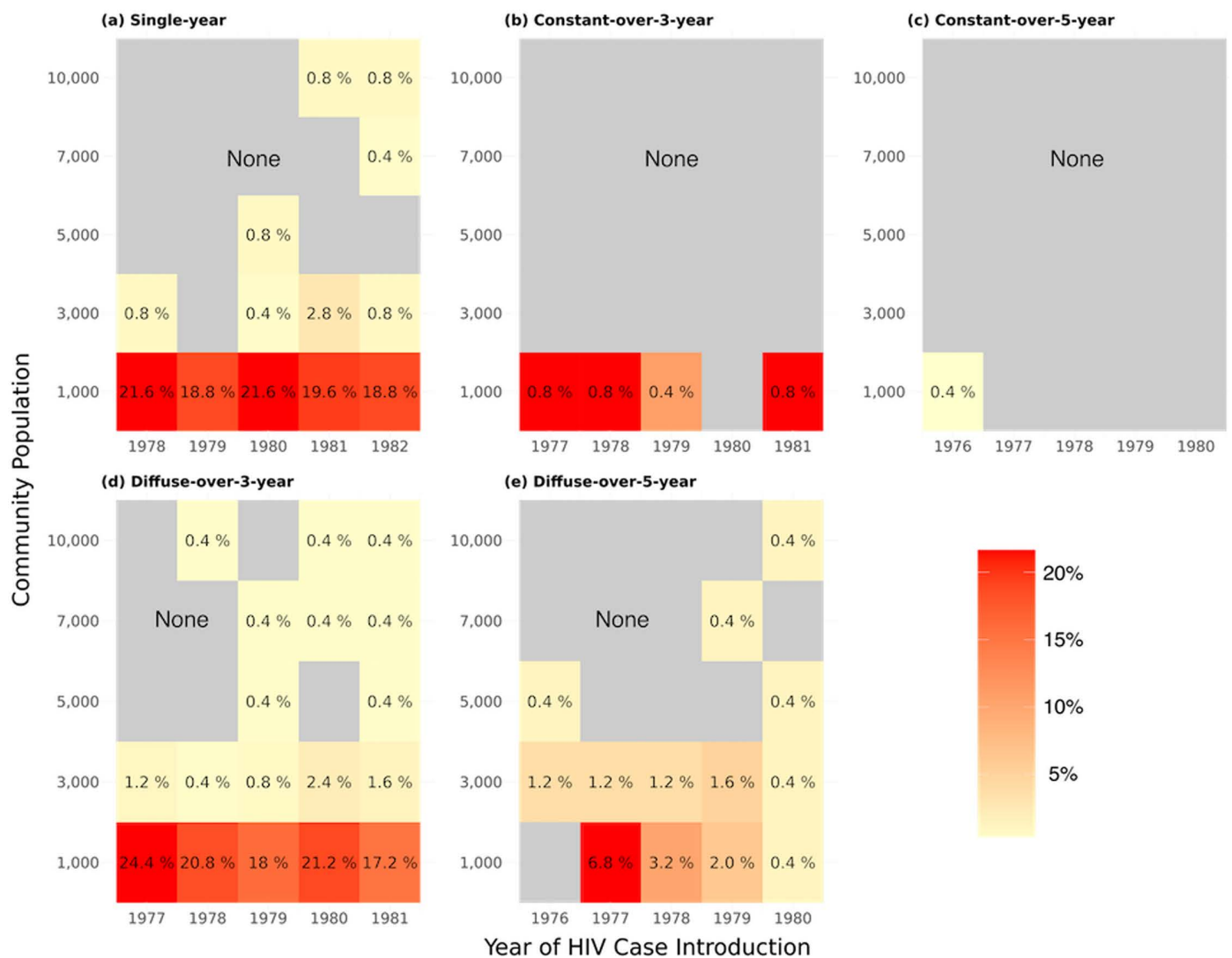


Fig 4. Proportion of communities identified as persistent prevalence outliers across community population and HIV case introduction pattern. The proportion of communities identified as persistent prevalence outliers for different HIV case introduction patterns and initial population sizes. The HIV case introduction patterns are: a: single-year (1 introduction event with probability p), b: constant-over-3-year (3 introduction events with p), c: constant-over-5-year (5 introduction events with p), d: diffuse-over-3-year (3 introduction events with $\frac{p}{3}$), and e: diffuse-over-5-year (5 introduction events with $\frac{p}{5}$). The columns represent the years of the first HIV introduction (1978, 1979, 1980, 1981, 1982). The rows correspond to the initial population sizes. Each cell in the heatmap indicates the proportion of communities identified as persistent prevalence outliers for a specific combination of HIV case introduction pattern, population size, and year of HIV case introduction.

<https://doi.org/10.1371/journal.pcbi.1013178.g004>

pre-specified threshold. DBSCAN results revealed a higher proportion of persistent outliers in smaller populations during 2005 and 2020, aligning with the threshold-based classification. This consistency suggests that stochastic effects play a substantial role in driving persistent prevalence outliers, regardless of the detection method used.

Sensitivity to alternative outlier definition thresholds

To further explore the robustness of our definition of outlier creation, we conducted a sensitivity analysis by varying the threshold value for defining outlier formation from 1.1 to 3 times the median HIV prevalence. When the threshold was set

Table 1. Autocorrelation of HIV prevalence across different time intervals and population sizes.

Population Size	Proportion of Prevalence Outlier (%)				
	1990 & 2005	1990 & 2020	1990 & 2050	2005 & 2020	2005 & 2050
1,000	42.4	30.8	31.6	32.0	32.0
3,000	22.4	0.4	8.4	0.4	6.8
5,000	12.4	0.0	4.0	0.0	4.0
7,000	10.0	0.0	2.0	0.0	1.6
10,000	4.4	0.0	2.0	0.0	1.6

The autocorrelation of prevalence is analyzed between the years 1990 and 2005, 1990 and 2020, 1990 and 2050, 2005 and 2020, as well as between 2020 and 2050, across initial populations of 1,000, 3,000, 5,000, 7,000, and 10,000 individuals. Scenarios with a low demographic risk profile and 1 seed in 1978.

<https://doi.org/10.1371/journal.pcbi.1013178.t001>

to 2 times the median prevalence, few outliers were detected across all simulation scenarios (S2 Fig). This observation was consistent across all initial population sizes and HIV case introduction patterns. When we further varied the persistent outlier definition threshold from 50% to 100% of the simulated time period, reducing the threshold to 50% of the time period allowed for persistent outliers to be detected for the population size of 1,000, but not for larger population sizes (S3 Fig).

Using DBSCAN, a higher proportion of simulations were identified as outliers in 2005 and 2020, which may reflect the epidemic dynamics during its expansion and stabilization phases, compared to 1990 and 2050 across all population sizes. Specifically, in 2005, 6.3% of the simulations across all population sizes were classified as outliers, while in 2020, this increased to 6.8%. During these years, community-level differences in transmission and intervention uptake could have amplified random fluctuations, leading to more communities deviating from the median prevalence. In contrast, the proportion of outliers in 1990 and 2050 was lower, with 1.6% and 0.1% of simulations, respectively. Smaller population sizes showed a higher propensity for outliers. In populations of 1,000 individuals, 23.8% of simulations were classified as outliers across the years, while in populations of 10,000 individuals, the outlier proportion dropped to 4.8%. This trend was consistent across all time points.

Discussion

This study explored the role of random fluctuations in the emergence and persistence of HIV prevalence outliers. We found that randomness can drive prevalence outlier formation, especially in smaller communities—a finding that was robust to alternative methods of outlier detection. In smaller communities, stochastic effects alone can result in the emergence and persistence of prevalence outliers, as random fluctuations in early case introductions disproportionately influence long-term prevalence trends. In contrast, larger communities typically require additional structural or behavioral factors—such as behavioral risks or gaps in prevention and treatment—for outliers to arise and persist. In smaller communities, limited sexual network connectivity and fewer external introductions can also reinforce high-prevalence states once established. Interestingly, the pattern and timing of HIV case importation—whether it occurred as an intense one-time event or diffused over several years—had minimal impact on outlier emergence and persistence. This suggests that once a community becomes a prevalence outlier due to stochastic effects, it may remain in a high-prevalence state over time, particularly in the absence of external mitigating factors. This persistence is likely driven by the long-term nature of HIV infection and the self-perpetuating cycle between high prevalence and high incidence.

The quantitative results reinforce this distinction between smaller and larger populations. In communities of at least 5,000 individuals, prevalence outliers defined as >1.5x the median prevalence were unlikely (<5%) to emerge due to randomness alone. Similarly, in communities of at least 3,000 individuals, the persistence of randomly formed prevalence outliers was rare. These findings suggest that while stochasticity plays a central role in smaller communities, larger

populations require additional epidemiological mechanisms—such as risk heterogeneity, mobility patterns, or differential access to prevention and treatment services—for outliers to form and persist. Although our model was calibrated to Kisumu County, the findings are not intended to be directly generalizable to all regions of sub-Saharan Africa. Differences in epidemiology, population structure, and healthcare access across settings may influence the extent to which stochastic processes contribute to prevalence outlier formation. Therefore, caution is warranted when extrapolating these results to other regions.

These findings contribute to a broader understanding of why prevalence outliers are formed and highlight the challenges in identifying their drivers. Similar to “patchy” epidemic patterns observed in other pandemics like COVID-19, where hotspots do not align neatly with known risk factors [46–48], HIV prevalence outliers in smaller communities may arise due to inherent stochasticity rather than specific, observable predisposing factors. This pattern has been observed in rural and peri-urban areas of Africa, where neighboring communities with comparable risk behaviors still experience striking differences in HIV burden [15–17]. Such spatial heterogeneity has been documented in high-prevalence geographies such as KwaZulu-Natal [37], Eswatini [46], and Lake Victoria basin [47], suggesting that stochastic effects could play a role in shaping prevalence outliers across multiple high-burden regions. Understanding how randomness contributes to these patterns is essential for refining epidemiological models and improving targeted intervention strategies.

However, while our results emphasize the role of randomness in prevalence outlier formation, they do not entirely dismiss the role of unobserved factors. In larger and more interconnected sexual networks, random fluctuations alone may be insufficient to explain the formation of prevalence outliers, suggesting that other factors—such as differential risk behaviors [30,33,34], social norms [29], or variations in access to healthcare [49]—may play a crucial role. Further research is needed to uncover these mechanisms in larger populations.

Community size—particularly when small—is a critical factor amplifying the influence of randomness, but other elements may also intensify these effects. Population mobility, social network connectivity, and localized health behaviors could introduce more complexity [50]. Small communities with closed or less interconnected sexual networks might be particularly susceptible to random variations, leading to concentrated outbreaks.

Recognizing randomness as a potential driver of prevalence outlier formation carries important implications for public health, particularly in HIV control efforts. One of the key challenges is that prevalence outliers driven by random fluctuations are inherently difficult to detect using traditional surveillance methods, which often rely on proxy indicators, such as behavioral risk factors or the presence of high-risk subpopulations [15,51]. Common approaches such as serosurveys (e.g., population-based surveys and ANC surveillance) and geospatial modeling of case data provide valuable insights [5,10,11,52], but they often require prohibitively large sample sizes to achieve the necessary granularity over large populations. As a result, these methods may overlook prevalence outliers that arise primarily due to stochastic effects [42,43] rather than identifiable risk factors, potentially leading high-prevalence areas undetected and missing opportunities for timely intervention. Addressing this challenge may require more spatially granular surveillance strategies [5,15], particularly in smaller communities. However, achieving this level of granularity over large populations requires prohibitively large sample sizes, making it challenging to implement at scale [53]. Serosurveys could help detect emerging hotspots early, even when they arise through stochastic processes [54]. Early detection is crucial, as it can prevent communities from becoming “trapped” in a high-prevalence state, a scenario that is particularly difficult to reverse given the long-term nature of HIV infection.

Our findings have broader implications for other infectious diseases with spatial clustering, such as tuberculosis, malaria, and COVID-19. Many of these diseases exhibit patchy incidence patterns, suggesting that stochastic processes could shape their spatial dynamics alongside biological, environmental, and behavioral drivers. For example, in the case of tuberculosis (TB), the transmission dynamics in high-burden settings are often influenced by local variations in population density, healthcare access, and socioeconomic conditions [55]. Our findings indicate that in smaller, more localized populations, the random introduction and spread of infection—amplified by pre-existing conditions like poor ventilation

or malnutrition—could create persistent clusters independent of these deterministic risk factors [56]. This highlights the importance of incorporating randomness into models and surveillance strategies for TB to better identify and target these clusters. In vector-borne diseases like malaria, random introduction of infected individuals or small-scale variability in vector density could lead to unpredictable hotspots [57]. Likewise, for rapidly spreading diseases like COVID-19, understanding that randomness can drive initial prevalence outlier formation underscores the need for flexible, adaptive surveillance systems capable of real-time detection [58] and response [59]. Applying adaptive surveillance strategies informed by these insights could enhance early detection and prevent widespread outbreaks by targeting resources and interventions to areas at risk of random epidemic emergence.

Overall, by highlighting the impact of randomness in the spatial dynamics of HIV, this study provides a framework for understanding how stochastic processes contribute to the emergence of prevalence outliers in HIV epidemics. Incorporating stochastic processes into epidemic models [60–62] and public health strategies will be essential for accurately predicting and managing the spatial patterns of these diseases, enabling more efficient use of resources and improving disease control and prevention efforts globally.

Future research should explore how ongoing and intermittent HIV case introductions, social network evolution, and varying access to healthcare services interact with stochastic processes to influence prevalence outlier formation in larger populations. Expanding the modeling framework to include genetic diversity in HIV strains [63], co-infections [64,65], and geographical migration patterns [37,52,66] could also provide a more comprehensive understanding of epidemic dynamics. Additionally, applying these approaches to other infectious diseases will help verify the broader applicability of these findings, ultimately improving our ability to predict, monitor, and control diverse outbreaks.

Our study has several important limitations. First, the modeling framework lacked explicit geographic demarcation of individual communities and inter-community migration. However, it did include a detailed sexual network. Social distance and geographic distance correlate, but social community sizes are ultimately impacted by multiple [64] factors including availability of transportation [15], labor migration patterns [67], social and cultural differences, and social norms around inter-cultural or inter-geographic sexual partnerships. Thus, interpretation of results through a geographic lens should be performed with caution and context-specific information.

Second, limited real-world data were available to validate our model, given that only a few HIV surveillance studies sampled participants densely enough to assess differences in HIV prevalence across small-scale communities and neighborhoods. Our findings were consistent with the magnitude of HIV prevalence differences observed in HIV surveillance sites in Africa [15–17], but given the limited number of sites that have undergone comprehensive HIV surveillance, we had insufficient data to formally validate our predictions.

Third, we focused exclusively on the initial importation of HIV cases in the early stages of the epidemic and did not consider the potential for new HIV importations occurring later in the epidemic. In reality, HIV is continually imported into communities—a process overlooked by our model and others [68–70]. However, as epidemics stabilize, the relative impact of migration-driven HIV introductions may diminish compared smaller relative to the size of the established epidemic. This pattern has been observed in other infectious disease settings where migration effects are most pronounced during early epidemic growth but play a reduced role once equilibrium is reached [71–74]. Nonetheless, even in stabilized epidemics, migration can still substantially contribute to incidence and prevalence, as highlighted in recent studies [75]. Future analysis could explore the effects of ongoing and intermittent HIV introductions throughout the epidemic.

Fourth, there is no widely-accepted quantitative threshold to define an HIV prevalence outlier. Our main analysis used an arbitrarily-chosen threshold of 1.5x the median of otherwise identical communities. However, in sensitivity analysis, we explored a wide range of alternative threshold definitions: 1.1x to 3x the median.

Finally, like all models, our model relies on assumptions in both structure and parameters. Parametrically, our model was calibrated to HIV epidemiology in Kisumu, Kenya, which may limit generalizability to other settings. Structurally, our model made numerous simplifications. For example, non-HIV STIs are modeled as a static risk rather than co-transmitting

on the sexual network; [76] we modeled only heterosexual and vertical HIV transmission (the primary transmission modalities in Eastern and Southern Africa); and certain biological details such as genomic differences in HIV virulence, transmissibility, and drug resistance are not included in the model [63]. However, the model chosen in this study is among the most detailed social network models available for HIV transmission in the African context and was particularly well-suited to study HIV prevalence outlier formation. Future research could revisit prevalence outlier analyses with alternative models and explore the amplification of random fluctuations in other infectious diseases beyond HIV.

Conclusions

Our study is the first, to our knowledge, to examine the role of random perturbations in the formation and persistence of HIV prevalence outliers. Smaller communities with relatively low rates of case importation appear to be relatively more prone to forming random HIV prevalence outliers than larger communities, with implications for HIV surveillance and for basic understanding of epidemic patterns in HIV and other infectious diseases. Recognizing that randomness can contribute to prevalence outlier formation underscores the value of adaptive public health strategies that consider stochastic processes alongside known epidemiological drivers. Integrating stochasticity into both surveillance and intervention efforts can lead to more resilient and adaptive responses, better suited to the complex and often unpredictable nature of epidemics. This shift in perspective may contribute to refining epidemic control strategies for HIV and other infectious diseases by emphasizing the importance of accounting for stochastic process.

Materials and methods

Model framework and calibration

We utilized the EMOD-HIV model, a previously validated and widely-used [60,77–80] agent-based HIV network transmission model developed by the Institute for Disease Modeling [61], to investigate the role of random fluctuations in the origins of HIV prevalence outliers. EMOD-HIV simulates the natural history of HIV infection, the effect of HIV treatment on health and transmissibility, and impact of HIV prevention, with care and prevention distributed according to a detailed care continuum [81].

As an agent-based model, EMOD-HIV allows for detailed representation of individual-level behaviors, relationship dynamics, and variations in sexual networks. HIV transmission is simulated on a sexual contact network, with different relationship types (in order from longest to shortest: marital, informal, transitory, and commercial) each having characteristic age-patterns of formation, duration, condom usage probability, and probability of openness to extra-relational sexual partners [82]. These elements are critical for understanding how random fluctuations in transmission can lead to varying epidemic outcomes. Additionally, EMOD-HIV simulates transmission at the level of individual coital acts, while accounting for factors such as condom use, antiretroviral therapy (ART) adherence, and sexually transmitted infections (STI) co-infections, makes it an ideal tool for exploring how small-scale random events and initial conditions can propagate to create larger spatial disparities in HIV prevalence. Vertical transmission from an infected mother to a child is also simulated, with transmission risk modified by the mother's use of effective ART or other antiretroviral medications for prevention of mother-to-child transmission (PMTCT). By allowing different patterns of case introductions and utilizing random number seeds to generate a wide range of potential epidemic trajectories, EMOD-HIV provides a robust platform for investigating how random perturbations in sexual networks and transmission dynamics may contribute to the emergence and persistence of HIV prevalence outliers.

EMOD-HIV was previously calibrated to HIV prevalence trends in a range of settings in Africa, including the Nyanza region of western Kenya, where it was validated by predicting HIV incidence in a large cluster-randomized trial measuring HIV incidence in sixteen communities [83]. These predictions were published while the trial was blinded, and then successfully validated against HIV incidence measurements. This western Kenya model has been used in several HIV

prevention modeling studies [60,77,84,85]. Likewise, Kisumu County was chosen as the illustrative example in this study due to its status as a well-documented HIV hotspot within the Nyanza region of western Kenya, an area that consistently reports some of the highest HIV prevalence rates in the country. This specific geography provides a relevant and representative setting for studying the role of randomness in prevalence outlier formation, as it combines high HIV burden with diverse community structures, varying population densities, and interconnected social networks.

The original model is subdivided into six geographic regions corresponding to the six counties of the Nyanza region: Homa Bay, Kisii, Kisumu, Migori, Nyamira, and Siaya Counties. For this study, we selected the Kisumu County geography as an illustrative example to investigate the emergence of outliers and assess whether they can arise from random fluctuations. Each simulation represents a single, independent community, meaning that no explicit spatial connections, inter-community migration, or geographic heterogeneity were included. The model focuses exclusively on within-community transmission, allowing us to isolate the role of within-community stochastic variation in shaping HIV prevalence patterns.

Simulation scenarios

To explore the factors contributing to the formation of HIV prevalence outliers, we simulated a total of 375 HIV epidemic scenarios (Table 2), varying the population size, timing of initial HIV case introduction, and the number of initially infected individuals. Each scenario was simulated with 250 different random number seeds to evaluate how stochastic fluctuations in early transmission dynamics impact the HIV epidemic. By maintaining all other epidemiological parameters constant while varying the random seed and initial seeding patterns, we specifically assessed the impact of randomness on the emergence and persistence of prevalence outliers. The 375 scenarios represented a range of initial population sizes (1,000; 3,000; 5,000; 7,000; and 10,000 individuals), timing of the start of the epidemic (1976–1984), duration during which initial HIV cases were distributed (1–5 years), and distribution of initial cases across behavioral risk groups. We

Table 2. Scenarios for HIV prevalence outlier emergence.

Population	Number of Seeds	Risk Level	Risk-Stratified Case Importation Fraction		Years of HIV Case Introduction					
1,000	1	High	0.0750		1978	1979	1980	1981	1982	
3,000		Medium	0.0180							
5,000		Low	0.0071							
7,000				p	$\frac{p}{3}$					
10,000										
	3	High	0.0750	0.0250	1977	1978	1979	1980	1981	
		Medium	0.0180	0.0060	1978	1979	1980	1981	1982	
		Low	0.0071	0.0024	1979	1980	1981	1982	1983	
			p	$\frac{p}{5}$	1976	1977	1978	1979	1980	
	5	High	0.0750	0.0150	1977	1978	1979	1980	1981	
		Medium	0.0180	0.0036	1978	1979	1980	1981	1982	
		Low	0.0071	0.0014	1979	1980	1981	1982	1983	
					1980	1981	1982	1983	1984	

The different simulation scenarios explored in this study. *Population* refers to the initial number of individuals in each simulated community. *Number of Seeds* represents the number of years over which initial HIV cases were introduced. *Risk Level* categorizes individuals based on their likelihood of acquiring HIV: High (7.5% of individuals seeded), Medium (1.8% seeded), and Low (0.7% seeded) if seeded at once. *Risk-Stratified Case Importation Fraction* indicates the proportion of individuals within each risk group who were initially seeded with HIV cases. The value p represents the seeding fraction for each group, while $\frac{p}{3}$ and $\frac{p}{5}$ correspond to scenarios where case introduction was spread over three or five years, respectively. *Years of HIV Case Introduction* denotes the years when initial infections were introduced into the population.

<https://doi.org/10.1371/journal.pcbi.1013178.t002>

then compared how random fluctuations in the HIV epidemic affected different scenarios and simulation runs, holding all other epidemiological assumptions constant.

Patterns of HIV case introduction

In epidemiological modeling, HIV-positive cases are introduced into the simulation to initialize the modeled epidemic in a process known as “seeding”. A random number seed is a value that initializes the random number generator used in the simulation, ensuring that each run has a unique sequence of random events, such as case introductions. In this study, we conducted comparisons across identical community configurations, where the only difference between simulations was the random number seed. This approach allows us to isolate the effects of stochastic variability and assess whether observed differences in HIV prevalence outliers could arise purely due to chance. Models generally only apply seeding early in a simulation to represent a net inflow and outflow (migration) of cases [62,68–70].

We modeled five different patterns of seeding, illustrated in Figs 5 and S4, each reflecting a chosen intensity and duration of case introduction into a community. In the single-year pattern, all cases (denoted as p) are introduced within a single year (e.g., 1978). In the diffuse-over-3-year pattern, introduction of cases is spread evenly over three years (e.g., from 1978 to 1980), with $\frac{p}{3}$ cases introduced per year, totaling p over the three-year period. Similarly, in the diffuse-over-5-year pattern, introduction of cases is spread evenly over five years (e.g., from 1976 to 1980), with $\frac{p}{5}$ cases introduced per year, summing up to p . In other scenarios, we allowed the number of introduced cases to rise in proportion to the seeding duration. In the constant-over-3-year pattern, p cases are introduced each year, so that a total of $3p$ cases are introduced. Similarly, in the constant-over-5-year pattern, a total of $5p$ cases are introduced. After this initial seeding period, no further HIV case importations occur in the simulations. All subsequent HIV transmission arises solely from within-community sexual networks, without external introductions. Initial seeding was conducted within three predefined risk groups (low, medium, and high risk), with a fixed proportion of individuals in each group assigned as initial cases. Specifically, $p = 7.5\%$ for the high-risk group, $p = 1.8\%$ for the medium-risk group, and $p = 0.7\%$ for the low-risk group. These proportions remained constant across all simulations to ensure comparability across scenarios.

Definition of HIV prevalence outlier existence and persistence

We defined HIV prevalence outlier community both cross-sectionally (outlier existence) and longitudinally (outlier persistence) based on deviations in HIV prevalence from the median of simulated communities under identical conditions. Outlier existence was initially defined by comparing HIV prevalence of the community at a given time (e.g., 1990) to the median prevalence across other simulation runs in the ensemble, which were simulated under identical assumptions and

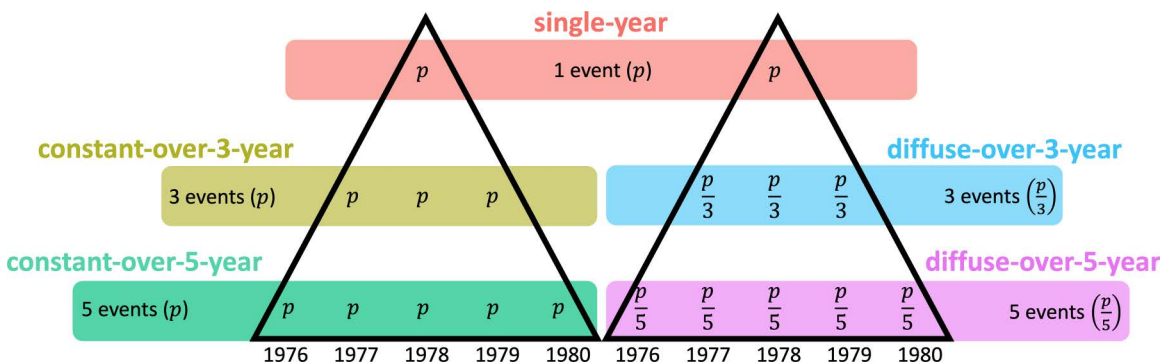


Fig 5. Overview of HIV case importation pattern in simulation scenarios.

<https://doi.org/10.1371/journal.pcbi.1013178.g005>

differed only due to stochastic variation. A community was classified as an outlier if its prevalence exceeded 1.5 times the median of communities configured with identical assumption. The choice of 1.5x the median was made to strike a balance between detecting substantial deviations in prevalence and obtaining a statistically robust sample of hotspots across a range of model assumptions regarding case importation. However, we recognize that this threshold is arbitrary, and to assess the robustness of our results, we conducted sensitivity analyses using alternative thresholds ranging from 1.1x to 3x the median prevalence.

Additionally, we applied Density-Based Spatial Clustering of Applications with Noise (DBSCAN) as an alternative method to detect outliers. DBSCAN identifies outliers based on local data point density, which allows for detecting replicates with prevalence rates that deviate from other replicates, regardless of the predetermined threshold.

To evaluate outlier persistence, we performed simulations spanning the period from epidemic seeding until 2050 in monthly time-steps. Simulations were classified as persistent outliers if they met criteria for outlier existence for $\geq 70\%$ of the simulated time-steps (137–149 time steps). The 70% threshold was selected to align with the time period from the center of seeding (1980) to the target of HIV epidemic control by 2030. This threshold was also varied in sensitivity analysis. DBSCAN was also applied to evaluate outlier persistence. To ensure robust detection of persistent outliers, we selected a threshold of 20 time steps. This value was chosen based on the observed distribution of outlier durations.

Statistical analysis

We used two-sample t-tests to compare HIV prevalence across different scenarios. To understand longitudinal patterns of outlier persistence, we calculated the autocorrelation of outlier status between specific pairs of observed years. Autocorrelation was calculated to assess the temporal persistence of outliers, providing insight into whether random prevalence outlier formation tends to persist over time. Autocorrelation was defined as the proportion of simulations identified as outliers in one year (e.g., 1990) that were also identified as outliers in a different year (e.g., 2005). This process was repeated for different year pairs. P-values were calculated with a null hypothesis of no correlation between prior and subsequent outlier status at the community level. Statistical analyses were performed using R version 4.0.3.

Supporting information

S1 Fig. Impact of population size, risk-stratified case importation fraction, and HIV case introduction pattern on the proportion of persistent HIV prevalence outliers.

(TIFF)

S2 Fig. Sensitivity analysis of HIV prevalence outlier proportions by population size, case introduction pattern, and observation year.

(TIFF)

S3 Fig. Sensitivity analysis of persistent HIV prevalence outliers by case introduction pattern, seeding year, and population size (A. high risk profile, B. medium risk profile, C. low risk profile).

(TIFF)

S4 Fig. Overview of HIV case importation pattern in simulation scenarios.

(TIFF)

S1 Table. Pairwise comparisons of mean HIV prevalence in 2018 by initial population size.

(XLSX)

S2 Table. Pearson correlation coefficients for HIV prevalence across various year pairs.

(XLSX)

Author contributions

Conceptualization: Nao Yamamoto, Daniel T Citron, Samuel M. Mwalili, Duncan K. Gathungu, Diego F. Cuadros, Anna Bershteyn.

Data curation: Nao Yamamoto.

Formal analysis: Nao Yamamoto.

Funding acquisition: Anna Bershteyn.

Methodology: Nao Yamamoto.

Software: Nao Yamamoto, Daniel T Citron.

Supervision: Daniel T Citron, Diego F. Cuadros, Anna Bershteyn.

Visualization: Nao Yamamoto.

Writing – original draft: Nao Yamamoto, Diego F. Cuadros, Anna Bershteyn.

Writing – review & editing: Nao Yamamoto, Daniel T Citron, Samuel M. Mwalili, Duncan K. Gathungu, Diego F. Cuadros, Anna Bershteyn.

References

1. Sovran S. Understanding culture and HIV/AIDS in sub-Saharan Africa. *SAHARA J.* 2013;10(1):32–41. <https://doi.org/10.1080/17290376.2013.807071> PMID: [23895330](https://pubmed.ncbi.nlm.nih.gov/23895330/)
2. Oster AM, Lyss SB, McClung RP, Watson M, Panneer N, Hernandez AL, et al. HIV Cluster and Outbreak Detection and Response: The Science and Experience. *Am J Prev Med.* 2021;61(5 Suppl 1):S130–42. <https://doi.org/10.1016/j.amepre.2021.05.029> PMID: [34686282](https://pubmed.ncbi.nlm.nih.gov/34686282/)
3. UNAIDS. Report – AIDS. 2020. <https://aids2020.unaids.org/report/>
4. United States Department of State. PEPFAR 2022 Country and Regional Operational Plan (COP/ROP) Guidance for All PEPFAR-Supported Countries. 2022. https://www.state.gov/wp-content/uploads/2022/02/COP22-Guidance-Final_508-Compliant-3.pdf
5. Achia T, Cervantes IF, Stupp P, Musingila P, Muthusi J, Waruru A, et al. Methods for conducting trends analysis: roadmap for comparing outcomes from three national HIV Population-based household surveys in Kenya (2007, 2012, and 2018). *BMC Public Health.* 2022;22(1):1337. <https://doi.org/10.1186/s12889-022-13633-8> PMID: [35831818](https://pubmed.ncbi.nlm.nih.gov/35831818/)
6. Nanyonjo G, Asiki G, Ssetaala A, Nakaweesa T, Wambuzi M, Nanvubya A, et al. Prevalence and correlates of HIV infection among adolescents and young people living in fishing populations along Lake Victoria Fishing Communities in Uganda. *Pan Afr Med J.* 2020;37:208. <https://doi.org/10.11604/pamj.2020.37.208.26124> PMID: [33505576](https://pubmed.ncbi.nlm.nih.gov/33505576/)
7. Mwanahapa P, Mtoro MJ, Gerald D, Horumpende P, Mujeeb S. Prevalence of HIV infection and uptake of HIV/AIDS services among fishermen on the shores of Lake Victoria in Kagera region, Northwestern Tanzania. *PLoS One.* 2025;20(1):e0315265. <https://doi.org/10.1371/journal.pone.0315265> PMID: [39854333](https://pubmed.ncbi.nlm.nih.gov/39854333/)
8. Kwena ZA, Njuguna SW, Ssetala A, Seeley J, Nielsen L, De Bont J, et al. HIV prevalence, spatial distribution and risk factors for HIV infection in the Kenyan fishing communities of Lake Victoria. *PLoS One.* 2019;14(3):e0214360. <https://doi.org/10.1371/journal.pone.0214360> PMID: [30908555](https://pubmed.ncbi.nlm.nih.gov/30908555/)
9. Kharsany ABM, Cawood C, Khanyile D, Lewis L, Grobler A, Puren A, et al. Community-based HIV prevalence in KwaZulu-Natal, South Africa: results of a cross-sectional household survey. *Lancet HIV.* 2018;5(8):e427–37. [https://doi.org/10.1016/S2352-3018\(18\)30104-8](https://doi.org/10.1016/S2352-3018(18)30104-8) PMID: [30021700](https://pubmed.ncbi.nlm.nih.gov/30021700/)
10. Tanser F, Bärnighausen T, Cooke GS, Newell M-L. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *Int J Epidemiol.* 2009;38(4):1008–16. <https://doi.org/10.1093/ije/dyp148> PMID: [19261659](https://pubmed.ncbi.nlm.nih.gov/19261659/)
11. Wand H, Ramjee G. Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. *J Int AIDS Soc.* 2010;13:41. <https://doi.org/10.1186/1758-2652-13-41> PMID: [20979651](https://pubmed.ncbi.nlm.nih.gov/20979651/)
12. Zulu LC, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994–2010. *BMC Infect Dis.* 2014;14:285. <https://doi.org/10.1186/1471-2334-14-285> PMID: [24886573](https://pubmed.ncbi.nlm.nih.gov/24886573/)
13. Waruru A, Achia TNO, Tobias JL, Ng'ang'a J, Mwangi M, Wamicwe J, et al. Finding Hidden HIV Clusters to Support Geographic-Oriented HIV Interventions in Kenya. *J Acquir Immune Defic Syndr.* 2018;78(2):144–54. <https://doi.org/10.1097/QAI.0000000000001652> PMID: [29474269](https://pubmed.ncbi.nlm.nih.gov/29474269/)
14. Tanser F, Bärnighausen T, Dobra A, Sartorius B. Identifying “corridors of HIV transmission” in a severely affected rural South African population: a case for a shift toward targeted prevention strategies. *Int J Epidemiol.* 2018;47(2):537–49. <https://doi.org/10.1093/ije/dyx257> PMID: [29300904](https://pubmed.ncbi.nlm.nih.gov/29300904/)
15. Tanser F, Hosegood V, Bärnighausen T, Herbst K, Nyirenda M, Muhwava W, et al. Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *Int J Epidemiol.* 2008;37(5):956–62. <https://doi.org/10.1093/ije/dym211> PMID: [17998242](https://pubmed.ncbi.nlm.nih.gov/17998242/)

16. Tanser F, Lesueur D, Solarsh G, Wilkinson D. HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. *Trop Med Int Health*. 2000;5(1):40–6. <https://doi.org/10.1046/j.1365-3156.2000.00513.x> PMID: [10672204](#)
17. Tanser FC, Kim H-Y, Mathenjwa T, Shahmanesh M, Seeley J, Matthews P, et al. Home-Based Intervention to Test and Start (HITS): a community-randomized controlled trial to increase HIV testing uptake among men in rural South Africa. *J Int AIDS Soc*. 2021;24(2):e25665. <https://doi.org/10.1002/jia2.25665> PMID: [33586911](#)
18. Lurie MN, Williams BG, Zuma K, Mkaya-Mwamburi D, Garnett GP, Sweat MD, et al. Who infects whom? HIV-1 concordance and discordance among migrant and non-migrant couples in South Africa. *AIDS*. 2003;17(15):2245–52. <https://doi.org/10.1097/00002030-200310170-00013> PMID: [14523282](#)
19. Kate Grabowski M, Lessler J, Bazaale J, Nabukalu D, Nankinga J, Nantume B, et al. Migration, hotspots, and dispersal of HIV infection in Rakai, Uganda. *Nat Commun*. 2020;11(1):976. <https://doi.org/10.1038/s41467-020-14636-y> PMID: [32080169](#)
20. Barnsley MF. *Fractals Everywhere*. Academic Press. 2014.
21. Barnsley MF. Fractal functions and interpolation. *Constr Approx*. 1986;2(1):303–29. <https://doi.org/10.1007/bf01893434>
22. Mandelbrot B. *Fractals and Chaos: The Mandelbrot Set and Beyond*. 2013.
23. Mandelbrot BB, Wheeler JA. The Fractal Geometry of Nature. *American Journal of Physics*. 1983;51(3):286–7. <https://doi.org/10.1119/1.13295>
24. Wallace R. A fractal model of HIV transmission on complex sociogeographic networks: towards analysis of large data sets. *Environ Plan A*. 1993;25(1):137–48. <https://doi.org/10.1068/a250137> PMID: [12285949](#)
25. Wallace R. A fractal model of HIV transmission on complex socio-geographic networks. Part 2: spread from a ghettoized “core group” into a “general population”. *Environ Plan A*. 1994;26(5):767–78. <https://doi.org/10.1068/a260767> PMID: [12287595](#)
26. Fauk NK, Ward PR, Hawke K, Mwanri L. Cultural and religious determinants of HIV transmission: A qualitative study with people living with HIV in Belu and Yogyakarta, Indonesia. *PLoS One*. 2021;16(11):e0257906. <https://doi.org/10.1371/journal.pone.0257906> PMID: [34780506](#)
27. Rujumba J, Kwiringira J. Interface of culture, insecurity and HIV and AIDS: Lessons from displaced communities in Pader District, Northern Uganda. *Confl Health*. 2010;4:18. <https://doi.org/10.1186/1752-1505-4-18> PMID: [21092165](#)
28. Lessler J, Azman AS, McKay HS, Moore SM. What is a Hotspot Anyway?. *Am J Trop Med Hyg*. 2017;96(6):1270–3. <https://doi.org/10.4269/ajtmh.16-0427> PMID: [28719289](#)
29. Lakew Y, Benedict S, Haile D. Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. *BMJ Open*. 2015;5(11):e008669. <https://doi.org/10.1136/bmjopen-2015-008669> PMID: [26589427](#)
30. Reniers G, Watkins S. Polygyny and the spread of HIV in sub-Saharan Africa: a case of benign concurrency. *AIDS*. 2010;24(2):299–307. <https://doi.org/10.1097/QAD.0b013e328333af03> PMID: [19890204](#)
31. Dietler D, Farnham A, Lyatuu I, Fink G, Winkler MS. Industrial mining and HIV risk: evidence from 39 mine openings across 16 countries in sub-Saharan Africa. *AIDS*. 2022;36(11):1573–81. <https://doi.org/10.1097/QAD.0000000000003294> PMID: [35730379](#)
32. Sagaon-Teyssier L, Balique H, Diallo F, Kalampalikis N, Mora M, Bourrelly M, et al. Prevalence of HIV at the Kokoyo informal gold mining site: what lies behind the glitter of gold with regard to HIV epidemics in Mali? A community-based approach (the ANRS-12339 Sanu Gundu cross-sectional survey). *BMJ Open*. 2017;7(8):e016558. <https://doi.org/10.1136/bmjopen-2017-016558> PMID: [28775190](#)
33. Delany-Moretlwe S, Bello B, Kinross P, Oliff M, Chersich M, Kleinschmidt I, et al. HIV prevalence and risk in long-distance truck drivers in South Africa: a national cross-sectional survey. *Int J STD AIDS*. 2014;25(6):428–38. <https://doi.org/10.1177/0956462413512803> PMID: [24352131](#)
34. Mlangeni N, Adetokunboh O, Lembani M, Malotle M, Nyasulu P. Access to HIV healthcare services by farm workers in sub-Saharan Africa (SSA): a systematic review protocol. *BMJ Open*. 2022;12(10):e059806. <https://doi.org/10.1136/bmjopen-2021-059806> PMID: [36307162](#)
35. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, et al. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *Lancet*. 2007;369(9562):643–56. [https://doi.org/10.1016/S0140-6736\(07\)60312-2](https://doi.org/10.1016/S0140-6736(07)60312-2) PMID: [17321310](#)
36. Akullian A, Onyango M, Klein D, Odhiambo J, Bershteyn A. Geographic coverage of male circumcision in western Kenya. *Medicine (Baltimore)*. 2017;96(2):e5885. <https://doi.org/10.1097/MD.0000000000005885> PMID: [28079830](#)
37. Ramjee G, Wand H. Geographical clustering of high risk sexual behaviors in “hot-spots” for HIV and sexually transmitted infections in Kwazulu-Natal, South Africa. *AIDS Behav*. 2014;18(2):317–22. <https://doi.org/10.1007/s10461-013-0578-x> PMID: [23934268](#)
38. Tsai AC, Hatcher AM, Bukusi EA, Weke E, Lemus Hufstедler L, Dworkin SL, et al. A Livelihood Intervention to Reduce the Stigma of HIV in Rural Kenya: Longitudinal Qualitative Study. *AIDS Behav*. 2017;21(1):248–60. <https://doi.org/10.1007/s10461-015-1285-6> PMID: [26767535](#)
39. Medema-Wijnveen JS, Onono M, Bukusi EA, Miller S, Cohen CR, Turan JM. How perceptions of HIV-related stigma affect decision-making regarding childbirth in rural Kenya. *PLoS One*. 2012;7(12):e51492. <https://doi.org/10.1371/journal.pone.0051492> PMID: [23251555](#)
40. Onu CC, Dworkin SL, Onger LG, Oyaro P, Neylan TC, Cohen CR, et al. Brief Report: Sexual Violence Against HIV-Positive Women in the Nyanza Region of Kenya: Is Condom Negotiation an Instigator?. *J Acquir Immune Defic Syndr*. 2017;74(1):52–5. <https://doi.org/10.1097/QAI.0000000000001135> PMID: [27509254](#)
41. Yebei VN, Fortenberry JD, Ayuku DO. Felt stigma among people living with HIV/AIDS in rural and urban Kenya. *Afr Health Sci*. 2008;8(2):97–102. PMID: [19357758](#)

42. Keeling MJ, Eames KTD. Networks and epidemic models. *J R Soc Interface*. 2005;2(4):295–307. <https://doi.org/10.1098/rsif.2005.0051> PMID: [16849187](https://pubmed.ncbi.nlm.nih.gov/16849187/)
43. Nàsell I. Stochastic models of some endemic infections. *Math Biosci*. 2002;179(1):1–19. [https://doi.org/10.1016/s0025-5564\(02\)00098-6](https://doi.org/10.1016/s0025-5564(02)00098-6) PMID: [12047919](https://pubmed.ncbi.nlm.nih.gov/12047919/)
44. Cuadros DF, Huang Q, Musuka G, Dzinamarira T, Moyo BK, Mpfu A, et al. Moving beyond hotspots of HIV prevalence to geospatial hotspots of UNAIDS 95-95-95 targets in sub-Saharan Africa. *Lancet HIV*. 2024;11(7):e479–88. [https://doi.org/10.1016/S2352-3018\(24\)00102-4](https://doi.org/10.1016/S2352-3018(24)00102-4) PMID: [38852597](https://pubmed.ncbi.nlm.nih.gov/38852597/)
45. Chang LW, Grabowski MK, Ssekubugu R, Nalugoda F, Kigozi G, Nantume B, et al. Heterogeneity of the HIV epidemic in agrarian, trading, and fishing communities in Rakai, Uganda: an observational epidemiological study. *Lancet HIV*. 2016;3(8):e388–96. [https://doi.org/10.1016/S2352-3018\(16\)30034-0](https://doi.org/10.1016/S2352-3018(16)30034-0) PMID: [27470029](https://pubmed.ncbi.nlm.nih.gov/27470029/)
46. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science*. 2020;369(6501):297–301. <https://doi.org/10.1126/science.abc1917> PMID: [32471856](https://pubmed.ncbi.nlm.nih.gov/32471856/)
47. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Appl Geogr*. 2020;118:102202. <https://doi.org/10.1016/j.apgeog.2020.102202> PMID: [32287518](https://pubmed.ncbi.nlm.nih.gov/32287518/)
48. Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. 2020;368(6490):493–7. <https://doi.org/10.1126/science.abb4218> PMID: [32213647](https://pubmed.ncbi.nlm.nih.gov/32213647/)
49. Cuadros DF, Li J, Mukandavire Z, Musuka GN, Branscum AJ, Sartorius B, et al. Towards UNAIDS Fast-Track goals: targeting priority geographic areas for HIV prevention and care in Zimbabwe. *AIDS*. 2019;33(2):305–14. <https://doi.org/10.1097/QAD.0000000000002052> PMID: [30557161](https://pubmed.ncbi.nlm.nih.gov/30557161/)
50. Rothenberg RB, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, Klovdahl AS. Social network dynamics and HIV transmission. *AIDS*. 1998;12(12):1529–36. <https://doi.org/10.1097/00002030-199812000-00016> PMID: [9727575](https://pubmed.ncbi.nlm.nih.gov/9727575/)
51. Tanser F, de Oliveira T, Maheu-Giroux M, Bärnighausen T. Concentrated HIV subepidemics in generalized epidemic settings. *Curr Opin HIV AIDS*. 2014;9(2):115–25. <https://doi.org/10.1097/COH.0000000000000034> PMID: [24356328](https://pubmed.ncbi.nlm.nih.gov/24356328/)
52. Cuadros DF, Li J, Branscum AJ, Akullian A, Jia P, Mziray EN, et al. Mapping the spatial variability of HIV infection in Sub-Saharan Africa: Effective information for localized HIV prevention and control. *Sci Rep*. 2017;7(1):9093. <https://doi.org/10.1038/s41598-017-09464-y> PMID: [28831171](https://pubmed.ncbi.nlm.nih.gov/28831171/)
53. Dwyer-Lindgren L, Cork MA, Sligar A, Steuben KM, Wilson KF, Provost NR, et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature*. 2019;570(7760):189–93. <https://doi.org/10.1038/s41586-019-1200-9> PMID: [31092927](https://pubmed.ncbi.nlm.nih.gov/31092927/)
54. Huerga H, Venables E, Ben-Farhat J, van Cutsem G, Ellman T, Kenyon C. Higher risk sexual behaviour is associated with unawareness of HIV-positivity and lack of viral suppression - implications for Treatment as Prevention. *Sci Rep*. 2017;7(1):16117. <https://doi.org/10.1038/s41598-017-16382-6> PMID: [29170407](https://pubmed.ncbi.nlm.nih.gov/29170407/)
55. Prado Junior JC, Medronho R de A. Spatial analysis of tuberculosis cure in primary care in Rio de Janeiro, Brazil. *BMC Public Health*. 2021;21(1):1841. <https://doi.org/10.1186/s12889-021-11834-1> PMID: [34641849](https://pubmed.ncbi.nlm.nih.gov/34641849/)
56. Li M, Guo M, Peng Y, Jiang Q, Xia L, Zhong S, et al. High proportion of tuberculosis transmission among social contacts in rural China: a 12-year prospective population-based genomic epidemiological study. *Cold Spring Harbor Laboratory*. 2022. <https://doi.org/10.1101/2022.04.18.22273772>
57. Grillet M-E, Barrera R, Martínez J-E, Berti J, Fortin M-J. Disentangling the effect of local and global spatial variation on a mosquito-borne infection in a neotropical heterogeneous environment. *Am J Trop Med Hyg*. 2010;82(2):194–201. <https://doi.org/10.4269/ajtmh.2010.09-0040> PMID: [20133991](https://pubmed.ncbi.nlm.nih.gov/20133991/)
58. Zhu S, Bukharin A, Xie L, Yamin K, Yang S, Keskinocak P, et al. Early Detection of COVID-19 Hotspots Using Spatio-Temporal Data. *IEEE J Sel Top Signal Process*. 2022;16(2):250–60. <https://doi.org/10.1109/jstsp.2022.3154972>
59. Base protocol for real time active random surveillance of coronavirus disease (COVID-19) - Adapting veterinary methodology to public health. <https://pubmed-ncbi-nlm-nih-gov.ezproxy.med.nyu.edu/32292815/> 2024 October 21.
60. Bershteyn A, Mutai KK, Akullian AN, Klein DJ, Jewell BL, Mwalili SM. The influence of mobility among high-risk populations on HIV transmission in Western Kenya. *Infect Dis Model*. 2018;3:97–106. <https://doi.org/10.1016/j.idm.2018.04.001> PMID: [30839863](https://pubmed.ncbi.nlm.nih.gov/30839863/)
61. Bershteyn A, Gerardin J, Bridenbecker D, Lorton CW, Bloedow J, Baker RS, et al. Implementation and applications of EMOD, an individual-based multi-disease modeling platform. *Pathog Dis*. 2018;76(5):fty059. <https://doi.org/10.1093/femspd/fty059> PMID: [29986020](https://pubmed.ncbi.nlm.nih.gov/29986020/)
62. Johnson L, Dorrington R. Thembe version 4.4: A model for evaluating the impact of HIV/AIDS in South Africa.
63. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S, et al. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science*. 2014;343(6177):1243727. <https://doi.org/10.1126/science.1243727> PMID: [24653038](https://pubmed.ncbi.nlm.nih.gov/24653038/)
64. Pawlowski A, Jansson M, Sköld M, Rottenberg ME, Källenius G. Tuberculosis and HIV co-infection. *PLoS Pathog*. 2012;8(2):e1002464. <https://doi.org/10.1371/journal.ppat.1002464> PMID: [22363214](https://pubmed.ncbi.nlm.nih.gov/22363214/)
65. Schindler S, Magalhães RA, Costa ELF, Brites C. Genital Cytokines in HIV/Human Papillomavirus Co-Infection: A Systematic Review. *AIDS Res Hum Retroviruses*. 2022;38(9):683–91. <https://doi.org/10.1089/AID.2021.0139> PMID: [35435746](https://pubmed.ncbi.nlm.nih.gov/35435746/)
66. Marty L, Cazein F, Panjo H, Pillonel J, Costagliola D, Supervie V, et al. Revealing geographical and population heterogeneity in HIV incidence, undiagnosed HIV prevalence and time to diagnosis to improve prevention and care: estimates for France. *J Int AIDS Soc*. 2018;21(3):e25100. <https://doi.org/10.1002/jia2.25100> PMID: [29603879](https://pubmed.ncbi.nlm.nih.gov/29603879/)

67. Lurie MN, Williams BG, Zuma K, Mkaya-Mwamburi D, Garnett G, Sturm AW, et al. The impact of migration on HIV-1 transmission in South Africa: a study of migrant and nonmigrant men and their partners. *Sex Transm Dis.* 2003;30(2):149–56. <https://doi.org/10.1097/00007435-200302000-00011> PMID: [12567174](https://pubmed.ncbi.nlm.nih.gov/12567174/)
68. Kerr CC, Stuart RM, Gray RT, Shattock AJ, Fraser-Hurt N, Benedikt C, et al. Optima: A Model for HIV Epidemic Analysis, Program Prioritization, and Resource Optimization. *J Acquir Immune Defic Syndr.* 2015;69(3):365–76. <https://doi.org/10.1097/QAI.0000000000000605> PMID: [25803164](https://pubmed.ncbi.nlm.nih.gov/25803164/)
69. Phillips AN, Cambiano V, Nakagawa F, Revill P, Jordan MR, Hallett TB, et al. Cost-effectiveness of public-health policy options in the presence of pretreatment NNRTI drug resistance in sub-Saharan Africa: a modelling study. *Lancet HIV.* 2018;5(3):e146–54. [https://doi.org/10.1016/S2352-3018\(17\)30190-X](https://doi.org/10.1016/S2352-3018(17)30190-X) PMID: [29174084](https://pubmed.ncbi.nlm.nih.gov/29174084/)
70. Stover J, Hallett TB, Wu Z, Warren M, Gopalappa C, Pretorius C, et al. How can we get close to zero? The potential contribution of biomedical prevention and the investment framework towards an effective response to HIV. *PLoS One.* 2014;9(11):e111956. <https://doi.org/10.1371/journal.pone.0111956> PMID: [25372770](https://pubmed.ncbi.nlm.nih.gov/25372770/)
71. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals* | Princeton University Press. 28 Oct 2007. <https://press.princeton.edu/books/hardcover/9780691116174/modeling-infectious-diseases-in-humans-and-animals> 22 Feb 2025.
72. WHO Ebola Response Team, Agua-Agum J, Allegranzi B, Ariyaratnam A, Aylward RB, Blake IM, et al. After Ebola in West Africa—Unpredictable Risks, Preventable Epidemics. *N Engl J Med.* 2016;375(6):587–96. <https://doi.org/10.1056/NEJMs1513109> PMID: [27509108](https://pubmed.ncbi.nlm.nih.gov/27509108/)
73. Stuckler D, Basu S, McKee M, Lurie M. Mining and risk of tuberculosis in sub-Saharan Africa. *Am J Public Health.* 2011;101(3):524–30. <https://doi.org/10.2105/AJPH.2009.175646> PMID: [20516372](https://pubmed.ncbi.nlm.nih.gov/20516372/)
74. Lurie MN, Stuckler D. The role of mining in the spread of TB in Africa: policy implications. *Expert Rev Anti Infect Ther.* 2010;8(11):1205–7. <https://doi.org/10.1586/eri.10.120> PMID: [21073282](https://pubmed.ncbi.nlm.nih.gov/21073282/)
75. Blenkinsop A, Monod M, van Sighem A, Pantazis N, Bezemer D, Op de Coul E, et al. Estimating the potential to prevent locally acquired HIV infections in a UNAIDS Fast-Track City, Amsterdam. *Elife.* 2022;11:e76487. <https://doi.org/10.7554/eLife.76487> PMID: [35920649](https://pubmed.ncbi.nlm.nih.gov/35920649/)
76. Boily M-C, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, et al. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis.* 2009;9(2):118–29. [https://doi.org/10.1016/S1473-3099\(09\)70021-0](https://doi.org/10.1016/S1473-3099(09)70021-0) PMID: [19179227](https://pubmed.ncbi.nlm.nih.gov/19179227/)
77. Bershteyn A, Sharma M, Akullian AN, Peebles K, Sarkar S, Braithwaite RS, et al. Impact along the HIV pre-exposure prophylaxis “cascade of prevention” in western Kenya: a mathematical modelling study. *J Int AIDS Soc.* 2020;23 Suppl 3(Suppl 3):e25527. <https://doi.org/10.1002/jia2.25527> PMID: [32602669](https://pubmed.ncbi.nlm.nih.gov/32602669/)
78. Bershteyn A, Jamieson L, Kim H-Y, Platais I, Milali MP, Mudimu E, et al. Transmission reduction, health benefits, and upper-bound costs of interventions to improve retention on antiretroviral therapy: a combined analysis of three mathematical models. *Lancet Glob Health.* 2022;10(9):e1298–306. [https://doi.org/10.1016/S2214-109X\(22\)00310-2](https://doi.org/10.1016/S2214-109X(22)00310-2) PMID: [35961353](https://pubmed.ncbi.nlm.nih.gov/35961353/)
79. Bansi-Matharu L, Mudimu E, Martin-Hughes R, Hamilton M, Johnson L, Ten Brink D, et al. Cost-effectiveness of voluntary medical male circumcision for HIV prevention across sub-Saharan Africa: results from five independent models. *Lancet Glob Health.* 2023;11(2):e244–55. [https://doi.org/10.1016/S2214-109X\(22\)00515-0](https://doi.org/10.1016/S2214-109X(22)00515-0) PMID: [36563699](https://pubmed.ncbi.nlm.nih.gov/36563699/)
80. Korenromp EL, Bershteyn A, Mudimu E, Weiner R, Bonecwe C, Loykissoonlal D, et al. The impact of the program for medical male circumcision on HIV in South Africa: analysis using three epidemiological models. *Gates Open Res.* 2021;5:15. <https://doi.org/10.12688/gatesopenres.13220.1> PMID: [33615145](https://pubmed.ncbi.nlm.nih.gov/33615145/)
81. Klein DJ, Bershteyn A, Eckhoff PA. Dropout and re-enrollment: implications for epidemiological projections of treatment programs. *AIDS.* 2014;28 Suppl 1:S47–59. <https://doi.org/10.1097/QAD.0000000000000081> PMID: [24468946](https://pubmed.ncbi.nlm.nih.gov/24468946/)
82. Bershteyn A, Klein DJ, Eckhoff PA. Age-dependent partnering and the HIV transmission chain: a microsimulation analysis. *J R Soc Interface.* 2013;10(88):20130613. <https://doi.org/10.1098/rsif.2013.0613> PMID: [23985734](https://pubmed.ncbi.nlm.nih.gov/23985734/)
83. Jewell BL, Balzer LB, Clark TD, Charlebois ED, Kwarisiima D, Kanya MR, et al. Predicting HIV Incidence in the SEARCH Trial: A Mathematical Modeling Study. *J Acquir Immune Defic Syndr.* 2021;87(4):1024–31. <https://doi.org/10.1097/QAI.0000000000002684> PMID: [33770065](https://pubmed.ncbi.nlm.nih.gov/33770065/)
84. Milali MP, Resar D, Kaftan D, Campbell J, Olowu A, Edwards D, et al. Cost-effectiveness of the dual prevention pill for contraception and HIV pre-exposure prophylaxis. *Front Reprod Health.* 2023;5:1144217. <https://doi.org/10.3389/frph.2023.1144217> PMID: [37266447](https://pubmed.ncbi.nlm.nih.gov/37266447/)
85. Mudimu E, Sardinia J, Momin S, Medina-Marino A, Bezuidenhout C, Bekker L-G, et al. Incremental costs of integrated PrEP provision and effective use counselling in community-based platforms for adolescent girls and young women in South Africa: an observational study. *J Int AIDS Soc.* 2022;25(2):e25875. <https://doi.org/10.1002/jia2.25875> PMID: [35129299](https://pubmed.ncbi.nlm.nih.gov/35129299/)