

Expanded Guidance on Collection, Use and Release of HIV Sequence Data

October 2021

Summary

- State and local HIV surveillance programs funded by CDC should collect HIV sequence data only in the form of Sanger sequences or, when next generation sequencing (NGS) has been conducted, consensus sequences. Currently, there is no documented public health benefit to collecting raw NGS data through HIV surveillance, and therefore, the risks of collecting these data outweigh any potential benefits.
- Analyses of HIV sequence data reported to HIV surveillance programs should not be interpreted as determining transmission direction or proving direct transmission between individuals, nor should analyses attempt to do so.
- CDC does not release HIV sequence data reported from HIV surveillance programs to GenBank or other publicly available sequence data repositories. State and local HIV surveillance programs funded by CDC and their academic partners should not release sequence data to GenBank or other publicly available sequence repositories without individual consent.

Justification

Collecting HIV sequence data is important for detecting and responding to HIV clusters and outbreaks. The sequences typically used for cluster detection by CDC and state and local health departments are either generated through Sanger sequencing or are consensus sequences generated after next generation sequencing (NGS). These sequences provide the most common nucleotide at each location in the sequenced gene(s) and do not provide the ability to identify direction of transmission. These data provide the necessary information to detect HIV clusters and outbreaks and respond with prevention and care interventions, which is the intended purpose for the collection of these data. Raw NGS HIV sequence data (i.e., the granular information about the sequence of each of the thousands of viruses in a person) may offer additional potential to infer direction of transmission between individuals, which is not the intended purpose of collecting these data. The collection of raw NGS data by HIV public health surveillance systems carries implications for the ethical balance of risks and harms of molecular sequence collection.

Currently, the risks of collecting raw NGS HIV data outweigh any potential benefits, as there is no documented public health benefit to collecting these data and these data may offer additional potential to infer direction of transmission. Perceived or actual ability to infer the direction of transmission could be associated with additional risks, such as release of data for non-public health purposes (i.e., for use in criminal or civil cases) or use of data to prompt public health actions that stigmatize people with HIV or people experiencing marginalizing circumstances. These factors could undermine community trust in this work.

Therefore, when laboratories conduct NGS, only HIV consensus sequences (and not raw NGS sequence data) should be reported to or collected by CDC-funded HIV surveillance programs.

Modernization of public health surveillance systems is important and should allow for flexibility in future needs, but raw NGS HIV data should not be collected in public health data systems at the current time. Continued reassessment of the relative benefits and risks of collecting these data, with guidance and input from key partners, including health departments, people with HIV, and laboratories, is essential to understand changing conditions that might shift the balance of benefit and risk. Conditions that might warrant reconsideration of the collection of NGS sequence data by public health agencies include: evidence demonstrating an established public health benefit to the collection and use of these data; people with HIV no longer experiencing disproportionate stigma or harms from criminal justice systems; community consultations to meaningfully involve people with HIV in this decision; and adequate protections in all jurisdictions that collect these data that ensure the data are used only for public health purposes by prohibiting data from being released or used as evidence in criminal or civil litigation.



Moreover, analyses of HIV sequence data reported to HIV surveillance programs, whether conducted by health departments or academic partners, should not be interpreted as determining transmission direction or proving direct transmission, nor should analyses attempt to do so. Rare exceptions might be considered (for example, when needed to establish or exclude an unusual route of transmission or when individual consent has been provided).

CDC does not release HIV sequence data reported from HIV surveillance programs to GenBank or other public repositories. Existing CDC guidance states that CDC-funded jurisdictions should not release identifiable, individual-level data to anyone outside of public health except in circumstances involving significant risk of harm to the public or if required by law. Even when required, only the minimum information should be released. **Therefore, HIV sequence data reported to CDC-funded HIV surveillance programs should not be released to GenBank or other public repositories by health departments or academic partners without individual consent.**