



SSuN Best Practice Note

Random Sample

PS24-0082 STI Surveillance Network Cycle 5 (SSuN)

This guidance applies to Strategy B activities

Rationale

Incorporating enhanced case-based surveillance (Strategy B) in SSuN Cycle 5 is motivated by the frequent occurrence of incomplete information about cases of sexually transmitted infections (STIs) regularly reported by healthcare providers and laboratories to public health authorities. Essential details, such as patient race, Hispanic ethnicity, sexual orientation, gender identity, and other pertinent clinical and behavioral characteristics are frequently omitted.

These data play a pivotal role in interpreting patterns in case reports and variations in disease burden, as well as in facilitating targeted efforts for improved prevention, control, and health equity. Carefully looking into a random group of reported cases of STIs and making sure it represents the whole population, facilitates the accurate estimate of cases that take into account important details about the people affected. This approach enhances our ability to inform effective public health strategies and interventions.

What is a Random Sample?

A simple random, or probability, sample is a selection of individuals from a larger population in which each member of the population has an equal chance of being chosen, allowing for a more accurate representation of the population as a whole. The utility of this sample for analysis lies in the fact that each selected member holds an equal 'probability' of being included in the sample. If this condition is met, the distribution of characteristics derived from investigating this sample closely mirrors what would be found if information could be obtained from the full population.



In the context of SSuN, in all jurisdictions participating in Strategy B:

- The full population is the total number of reported gonorrhea **and** early syphilis cases;
- The random sample is a representative subset of all reported gonorrhea and early syphilis cases (more information on how to select the random sample can be found on p.13-14 of the SSuN Cycle 5 NOFO);
- The ratio of sample cases to total reported cases is termed the **sample fraction**; and
- The proportion of successfully investigated cases among the sampled ones is referred to as the **effective sample**.

Ensuring there are no **intentional or inadvertent** biases in the sample is crucial, meaning that there should be no restrictions on the selection of reported cases for sampling. The only allowable exception is that the sample size may vary by geographic unit by design.

It is imperative that all cases, irrespective of characteristics such as patient age, gender, race, Hispanic ethnicity, source of report, or method of data entry (e.g., electronic versus hand-entered), have an equal probability of being sampled. While establishing a different sample size based on predefined geographic criteria (such as the patient's county of residence) is acceptable, this can be adjusted in the analysis by creating design weights. However, within the geographic unit, no other criteria should be used to adjust the sample size.

Considerations for Implementation

Recommended Approach for Random Sample Selection

The recommended approach for random sample selection is to use the built-in random number functions in your STI surveillance information system. Popular software like Oracle or SQL provide these functions, accessible through stored procedures or custom programming.

Step 1. Ensure that the random number function generates a value between 0 and 1.0 (or 0 and 100, depending on the function). This range should remain constant throughout the sampling process.

- Create a dedicated container (e.g., a variable or separate column in data tables) to store the result of the random number generator permanently. Run and capture the random number only once for each record.

Step 2. Select a sample by comparing the randomly generated number to a fixed range (sampling fraction).

- Records falling within this range are considered to be 'in the sample.'
- Execute this step once, setting a binary indicator or flag to mark the case as 'in the sample.'
 - This step adds flexibility, allowing for different sample fractions based on diseases, counties, regions, or areas.
- Build a flexible system that uses the randomly generated number for each record to customize case sampling based on your local project design.
 - For instance, you can adjust sample size:
 - By area
 - By disease
 - Over time to adapt to changing conditions



Automating the generation of a random number and establishing a versatile method to adjust and modify your case sample size based on this random number can serve as a powerful tool for evaluating and enhancing surveillance activities across various disease and data quality issues.

Other Resources

- ❖ SSuN website: <https://www.cdc.gov/std/ssun/default.htm>

CDC's Division of STD Prevention, Enhanced Surveillance and Special Studies Team created this series of documents for CDC staff working on PS24-0082 STI Surveillance Network (SSuN) and for applicants and recipients of that NOFO, to help clarify strategies outlined there and to support project implementation.