Enhanced Gonorrhea Surveillance Toolkit

Council of State and Territorial Epidemiologists

January 2020

CSTE

Acknowledgements

This Enhanced Gonorrhea Toolkit was made possible by the Council of State and Territorial Epidemiologists (CSTE), in partnership with the Centers for Disease Control and Prevention (CDC) and Cardea.

CSTE

Ashley Vineyard, MPH

CARDEA
Training, Organizational Development and Research

Michelle Daugherty
Molly Feder, MPH
Wendy Nakatsukasa-Ono, MPH
Eric Wheeler
Amanda Winters, MPA, MPH

CDC
Centers for Disease Control and Prevention

Cassandra Davis, MPH
Mark Stenger, MA
Elizabeth Torrone, MSPH, PhD
Emily Weston, MPH
Hillard Weinstock, MD, MPH

Many thanks to those who created and appeared in the video segments…

CDPH California Department of PublicHealth

Nicole Burghardt, MPH
Emily Han, MPH

NYC Health

Ellen Klinger, MPH

Department of Public Health
City of Philadelphia

Robbie Madera, MPH

In addition to those listed above, thank you to those who contributed to the development of this toolkit, including STD Surveillance Network (SSuN) collaborators and to state and local STD prevention and control programs across the U.S. We are grateful for your support.

Table of Contents

Note: This Toolkit is primarily meant to be used as an electronic document. If you wish to print this document, use 11" x 17" (tabloid) paper. If printed, you will not have access to links or other media embedded in this document.

SECTION 1

Getting Started

Who is this toolkit for?

Project areas conducting enhanced GC surveillance, including:

- Strengthening STD Prevention and Control for Health Departments (STD PCHD) recipients seeking additional guidance on expectations for enhanced gonorrhea surveillance
- Those with no enhanced surveillance background who desire guidance
- Those who already conduct enhanced gonorrhea surveillance but desire additional guidance

Reminder

STD PCHD includes an enhanced gonorrhea surveillance activity to help project areas:

- Strengthen surveillance capacity
- Better understand the epidemiology of gonorrhea in their project area
- Use local data to inform prevention and control opportunities

KEY TERMS

Gonorrhea (GC)
Gonorrhea is a sexually transmitted infection (STI) that can infect both men and women. It most commonly causes infections in the genitals, rectum, and throat. Infections can cause symptoms such as painful urination, discharge, and sore throat; however, many people with gonorrhea do not have any signs or symptoms. In the United States, gonorrhea is a very common STI, especially among young people ages 15–24 years.

Case-based GC surveillance
Laboratory reports indicating a gonococcal infection are sent to state (or local) health departments by public and private laboratories. In some project areas, diagnosing providers also send case report forms for diagnosed infections to the health department. Health departments review laboratory and provider reports to identify cases that meet the CSTE case definition for gonorrhea and de-duplicate cases if necessary. Case report data are sent to CDC at least weekly through the National Notifiable Disease Surveillance System (NNDSS). Case data are used to describe local and national trends in diagnosed infections.

Enhanced case-based GC surveillance
Additional surveillance activities that expand upon routine case-based GC surveillance efforts, including collection of data beyond those that have been reported by providers and/or laboratories to state and local health agencies. This includes conducting provider follow-up and brief patient interviews. As the number of GC cases in a project area may prohibit enhanced investigation of all reported cases, enhanced surveillance can be completed on a representative, random sample of GC cases from a well-defined high morbidity area or a project area as a whole.

What does enhanced GC surveillance tell us that routine GC surveillance doesn't?

Given the burden of reported GC in most project areas, it is unlikely that all cases are investigated and as a result, most case reports only include the limited information provided on the laboratory report (ie: sex, age, diagnosing provider, specimen collection date, and county of residence). Enhanced surveillance can help gather other important demographic, clinical, and behavioral variables, such as race/ethnicity, HIV status, and gender of sex partners, which could help better describe the GC epidemic and inform prevention strategies.

SECTION 1: GETTING STARTED

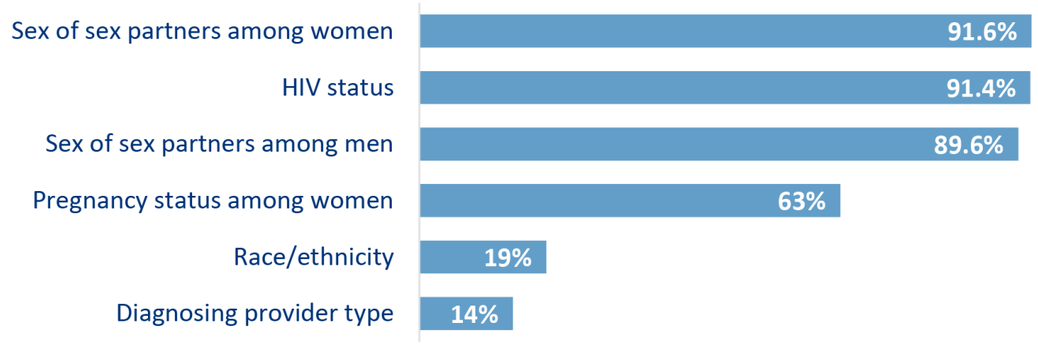
Enhanced GC surveillance:

1 Includes data that are not routinely collected, including (but not limited to):

- Demographic and clinical information, such as gender and number of sex partners, HIV and pregnancy status
- Treatment information, such as name and dose of antibiotics
- Sequelae, such as pelvic inflammatory disease (PID), disseminated gonococcal infection (DGI)
- STI co-infection

2 Ensures that core variables are collected and complete for cases targeted for enhanced investigation.

Percentage (%) of national gonorrhea case reports to CDC with missing data for select variables, 2017



What are the benefits of enhanced surveillance?

With this additional information, a project area can better:

1 Monitor and interpret trends in reported cases of GC

2 Assess inequalities in the burden of disease by population characteristics

3 Understand STI-related care seeking behaviors

4 Respond with appropriate and targeted prevention, screening and treatment interventions

Reminder

Engage Stakeholders

Engage local project areas in each step of enhanced surveillance to ensure buy-in, efficiency, and successful program outcomes.

Consider who key stakeholders are and how they should be engaged in each step of enhanced surveillance.

Video: Engaging Stakeholders for Enhanced GC Surveillance: Examples from California

https://vimeo.com/372463046

TOOLS

- Enhanced STD Surveillance Network (eSSuN) Protocol and Project Implementation Guide
- Gonorrhea — CDC Fact Sheet
- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- Technical Assistance Note # 2 | Enhanced surveillance for GC cases

SECTION 2

Defining your Surveillance Population for Enhanced GC Surveillance

Who is in my population?

When deciding how to define your surveillance population for enhanced GC surveillance, you will need to think about…

WHEN

Enhanced surveillance does not need to be conducted throughout the entire year and it is possible for a project area to select a pre-defined and specific time period for the activity.

How do I define a "specific period of time"?

The specific period of time you select will vary across project areas. It may be a full year or it may be part of year, such as six months. It is generally recommended to collect data for a period of three months or more to ensure data are reliable.

WHERE

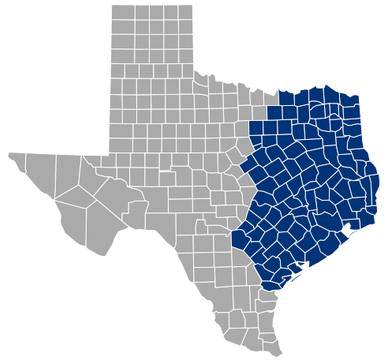
What is the geographic area of your surveillance population?

This area should be well-defined

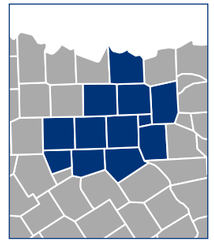
How can you tell if you have selected a "well-defined geographic area"?

A well-defined geographic area could be a:

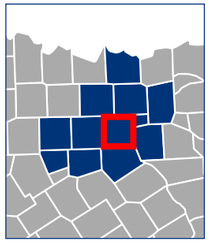
State (Texas)



Cluster of counties (northeastern Texas)



Large metropolitan area (Dallas-Fort Worth)



Single county or city (Dallas)

This area…

- May be dependent on where most cases of GC are from
- May be dependent on where there are high incidence or high re-infection rates
- Could include areas of interest for a project areas (e.g., border communities, college towns, high proportion of international travelers)
- Should have a quantifiable population so that you can calculate rates

The size of the area may differ based on morbidity

What is a high, middle, or low morbidity area?

A "high morbidity area" generally refers to the location from which most cases are reported. This could be an entire state, but more often it is a few counties in a region or one metropolitan area or city. A "low morbidity area" refers to a location from which few cases are reported, and a "middle morbidity area" would be somewhere in between a high and low morbidity area. CDC does not have standardized cutoffs to define morbidity levels.

EXAMPLE: In lower morbidity areas, you might select the entire project area because it may be feasible to conduct enhanced GC surveillance for this full population. In higher morbidity areas, you might need to select a cluster of counties or a large metropolitan area.

What if GC epidemiology varies in your project area? Will the estimates from just one high or low morbidity area be biased?

As long as you do not try to generalize the findings from one morbidity area beyond that specific area, your estimates should not be biased. Your findings are meant to inform your understanding of the GC epidemic in the area you select. You have the option of sampling cases from your entire project area. However, starting with a well-defined area can help you define your methods, which you can later expand to other areas.

Recommendation
Use data you are currently collecting to understand the landscape and inform the region you select for enhanced surveillance.

SECTION 2: SURVEILLANCE POPULATION

WHO

Which cases will be included in your surveillance population under enhanced surveillance?

- If feasible, all cases in the selected geographic area for the identified time period should be included in enhanced investigation.
- If enhanced investigation of all cases is not feasible, then a representative sample of all cases in the selected geographic area for the identified time period should be included in enhanced investigation (see Section 3 for how to identify a representative sample).

TOOLS

- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar

SECTION 3

Random Sampling

How do I ensure my sample is representative of all reported gonorrhea cases?

What is a random sample?

A random sample is a subset of cases picked at random from all reported cases. A true random sample ensures each reported case has the same probability of getting picked for the sample. When this is true, the distribution of characteristics in this sample of cases should closely resemble that of all reported cases in a project area. The ratio of sampled cases to all reported cases is known as the sample fraction.

When should you select a random sample?

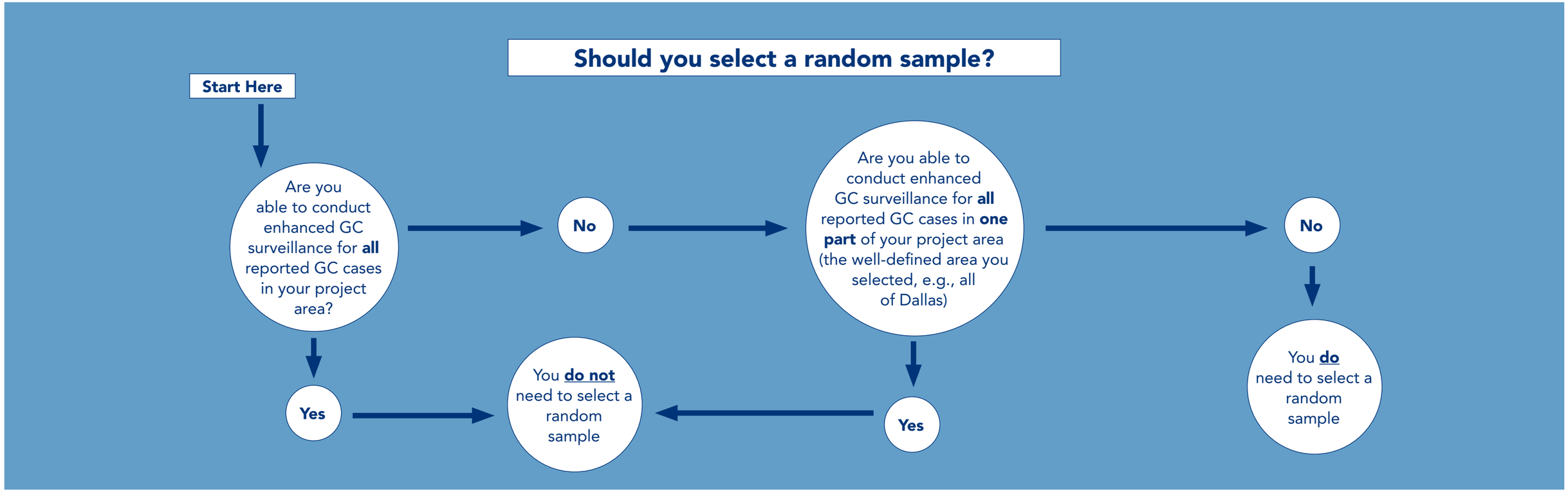
If you are able and plan to conduct enhanced surveillance on all reported cases in the well-defined, high morbidity area you select, then you do not need to take a sample. Otherwise, you will need to take a random sample of your reported GC cases to make sure the enhanced surveillance you conduct is representative.

Best practice is to choose cases as close as possible to the initial report and to capture the result permanently in the electronic case record.

Recommendation

A best practice is to conduct random sampling at the state level, even if you are only going to focus on one geographic area in the state. For example, Missouri would identify a random sample from the entire state (and identify those randomly selected cases with a 'yes' for the case sampled variable) even if they are only planning to conduct enhanced investigation in Kansas City. This would allow the project area to make changes to the selected geographic area with minimal changes to the sampling methodology (i.e., expand to conduct enhanced surveillance in both Kansas City and Branson).

Note: A city or county that receives separate STD PCHD funding (e.g., Los Angeles, New York City, Philadelphia, etc.) should draw their random sample from their own geographic area under surveillance.



Why is random sampling needed for enhanced GC surveillance?

Most project areas do not have the resources to investigate all of their cases. A random sample allows you to select a smaller number of your cases to investigate, while ensuring that the sample is representative of the population you are hoping to learn more about, i.e., to prevent biases.

Which cases should be included in your random sample?

Your random sample must be **selected from ALL GC cases** in your selected area during the time period of interest after de-duplication.

In other words, there should be no restrictions on which reported cases get sampled if you wish to have your sample reflect the full population. All reported gonorrhea cases in your pre-defined surveillance population, regardless of any characteristic such as patient gender, race, ethnicity, age, source of report or method of data entry, must have the same probability of being selected.

For example, conducting enhanced investigations only among cases diagnosed in STI clinics or only among young women aged 15–24 years would not be representative of all cases in the area.

The only exception to this rule is that the sample size may be allowed to differ by pre-defined geographic criteria (such as patient's county of residence) because this can be adjusted for in analysis by creating design weights. However, within the geographic unit, no other criteria should be used to adjust the sample size.

If surveillance is carried out for a relatively short period of time (i.e., three months) paper reports should be entered quickly to make sure that all cases are available in the electronic surveillance system and able to be sampled.

REVIEW

What is bias?

Bias is systematic error that leads to an incorrect estimate of an association between exposure and risk of disease.

If your sample is biased, it could lead to an incorrect interpretation of your results. For example, if your cases are only coming from STI clinics, which may have a higher incidence of GC than other types of clinics, your findings will not be representative of the full project area.

SECTION 3: RANDOM SAMPLING

How do you obtain a random sample of your cases?

BEFORE SAMPLING

Prior to obtaining your sample, consider the number or percent of total cases you wish to sample. This number will differ for each project area based on the response rate assumptions. In general, the minimum percent of total reported GC cases that should have complete investigations conducted is 5% because a smaller sample size may lead to unstable estimates or wide confidence intervals. However, a larger sample is recommended because the sampling fraction must also account for non-response. For example, assuming non-response is 50% of individuals, 10% is the minimum percent of cases that should be randomly sampled.

SITE EXAMPLE

Pre-defined area selected	3-county metro area
Time period selected	6 months
Number of reported cases in area during time period	5,000
Sample percent desired	5%
Goal	250 completed investigations (5,000 X 0.05)
Response rate assumption	50%
Number of cases needed in random sample	500 cases (250 / 0.50)
Sample fraction	10% (500/5,000)

The number of cases that makes sense to sample will vary across project areas and depend on the project area's…

It is also important that project areas have the infrastructure in place to select a representative random sample from ALL reported cases of GC in the well-defined geographic area prior to implementing patient and provider investigations. Selecting a valid, representative sample may take time, especially if the project area has not performed enhanced surveillance before.

Reminder

- The number of cases will vary across project areas
- Try to aim for complete investigations for at least 5% of all of the morbidity in the selected area
- Be aware that small sample sizes may lead to unstable estimates or wide confidence intervals
- If you have questions about generating your random sample, you can reach out to your CDC project officer

SECTION 3: RANDOM SAMPLING

Eligibility requirements

You may determine that some cases are not eligible for enhanced investigation due to local or other policies that are specified prior to initiating sampling. Eligibility criteria should be clearly defined and if cases are not eligible for inclusion, the reason should be recorded. Reasons for excluding a case might include:

Reminder

If you determine that some cases are not eligible for inclusion in your sample based on your selected criteria, your random sample will only be representative of that group rather than all reported GC cases.

DURING SAMPLING

Generating a random sample

Sampling methodology should allow you to modify the sample size, if needed, based on your success in conducting investigations to make sure you are able to complete your target number of investigations.

STEP 1: Obtain the records of reported GC cases from which the random sample will be selected

The random sample should come from all cases of laboratory confirmed gonorrhea received by health departments from all public and private sources included in your pre-defined surveillance population. Gonorrhea 'reports' will include provider case reports, laboratory reporting, and/or any other original source documents as appropriate given the specific surveillance infrastructure in the project area.

STEP 2: Consider your sampling methodology

Sampling methodologies will vary across project areas. However, to ensure the quality of the random sample, project areas should follow certain key criteria:

- All gonorrhea cases reported to the state or local health department during the selected surveillance time period and from the selected geographic area should have the same chance of being sampled.
  - If possible, modify your surveillance data management system to select the random sample automatically as cases are entered into the system instead of using an external batch process
  - If you must obtain your sample through a batch process, the process for sampling cases should be timely, with no more than 15 calendar days elapsing between receipt of the record at the health department, inclusion in a sample frame and subsequent referral for enhanced investigation.
- Records sampled should be referred for enhanced investigation as soon as possible

Recommendation

You may wish to stratify your sample by county or another geography-based factor to balance work-load within collaborating project areas. However, do not stratify based on other demographics, such as patient sex, age, race, ethnicity or provider characteristics, as this will prevent your sample from being random.

SECTION 3: RANDOM SAMPLING

STEP 3: Pick a random sample of your cases

The best practice is to select your sample as close as possible to the time you receive the initial case report and to capture the result permanently in the electronic case record:

- Use the tools you have
  - Data management software (e.g., Oracle, SQL, etc) have random number functions built into their systems that can be accessed through stored procedures or custom programming…use them!
  - The function must randomly generate a number between 0 and 1.0, or between 0 and 100, depending on the function you use
  - This range should not be altered or changed once sampling has begun
- Create a variable or separate column in your data tables where the result of the random number generator can be stored
- Make sure that this random number is only run and captured ONCE for any given record.
- Select a fixed cut-off between your random numbers that are generated from your random number generator. For example, if each case is assigned a number between 0 and 100 and your sample fraction is 20%, then any number less than 20 would be included in your sample and any number greater than 20 would not be included in your sample.
- Create another variable or separate column in your data tables that is binary (e.g., 0, 1) or in some way flags those cases as being in your sample

Recommendation

Consider flexible ways to use the random number generated for each record so that cases can be sampled differently depending on your specific project. For example, you may want sample size to increase or decrease based on area/county, disease, time period, etc. You also want to be sure that enough records are included in the random sample so that you will have enough completed case investigations to fulfil your project goals.

AFTER SAMPLING

After obtaining your random sample, check to make sure that:

- The sample is similar to all reported gonorrhea cases within your identified surveillance population
  - In a true random sample, the cases that are selected for enhanced GC surveillance should be similar to ALL reported GC cases in regards to their distribution of key characteristics, such as sex, age, and geographic area when more than one county or region is included.
- There are no biases to your results due to non-response
  - Your team should monitor those who decline or are not reached for interviews to make sure that there is not something different about that group compared with the rest of the sampled cases that makes them more or less likely to complete an investigation. For instance, is there a specific age group, sex, race/ethnicity that is not responding?
  - Although response rates sometimes vary, it is important to try and identify response bias so that you can address it, if possible

Recommendation

Your sample should be monitored against all eligible GC cases regularly (monthly or quarterly) for representativeness to be sure that any accidental biases are not affecting your sample, and thus the conclusions you draw from the data. As a general rule, small variations (i.e., less than 2%) in the distribution of key demographics are expected, especially when there are few GC cases. As more GC cases are reported, the differences should become smaller. If they become larger, or if the differences are more than 2% between you sample and the overall pool of GC cases, you will need to check your process, sample fraction, and number generator to make sure that all steps are functioning properly.

SECTION 3: RANDOM SAMPLING

What should you do if non-response varies between groups?

Creating and applying weights that account for non-response can minimize bias due to differential response rates. Creating non-response weights is beyond the scope of this tool-kit, but see Section 6: Weighting data for additional tools and resources.

TOOLS

- Maryland example SAS code for case selection
- Michigan example SAS code for sampling
- SSuN Best Practice Note — Strategy B: Random Sampling
- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology

SECTION 4

Collecting Data for Enhanced Gonorrhea Surveillance

What data should I collect and how?

BEFORE DATA COLLECTION

What data do you already collect?

Before you begin data collection for enhanced surveillance, review the data that are already collected and think about the questions you will be able to answer already. Ask yourself:

1. Are my existing data representative of all gonorrhea cases?
2. How complete are the data that I have?

DURING DATA COLLECTION

What data should be collected?

Data will include clinical, demographic, and provider-level variables as described and listed in CDC's STD PCHD technical assistance materials. Some variables are considered "**core**" and some " **optional**". Core variables are required to collect, whereas optional variables are recommended but not required.

Core Variables

- ☑ Age (date of birth)*
- ☑ Sex*
- ☑ County*
- ☑ Diagnosing facility type* (e.g., STI clinic, correctional facility)
- ☑ Specimen collection date*
- ☑ All anatomic site(s) of infection*
- ☑ Case sample**
- ☑ Race/ethnicity
- ☑ Gender identity/sexual orientation
- ☑ Clinical Symptoms and signs (healthcare seeking behaviors)
  - Length of time symptoms were present
- ☑ Sex of sex partner(s)
- ☑ Pregnancy status
- ☑ HIV status
- ☑ Previous history of GC infection
- ☑ Gonorrhea-related sequelae (PID, disseminated gonococcal infection)
- ☑ Date of diagnosis
  Note: *if the person had a second infection after 30 days of an infection, they would be considered a new case*
- ☑ Treatment provided (name and dose of treatment/antibiotic)
- ☑ Date of treatment
- ☑ Co-infection with other STIs
- ☑ History of substance abuse (IVDU, etc.)
- ☑ Partner treatment (EPT provision)

* These variables are considered core for GC surveillance (STD PCHD surveillance component 2a). These variables, along with the others listed in this table, are all considered core for enhanced surveillance (STD PCHD surveillance component 2b).

** Case sample is a new variable that was added to NETSS that allows you to note whether the case was randomly sampled for enhanced GC surveillance.

Optional Variables

- ☑ Insurance status/coverage
- ☑ Antibiotic use in the last 2 weeks, including the name of the antibiotic (if possible)
- ☑ Travel and related sexual history in past 30 days
  - Foreign vs. domestic travel
  - Country of birth of partner

Sexual partner characteristics

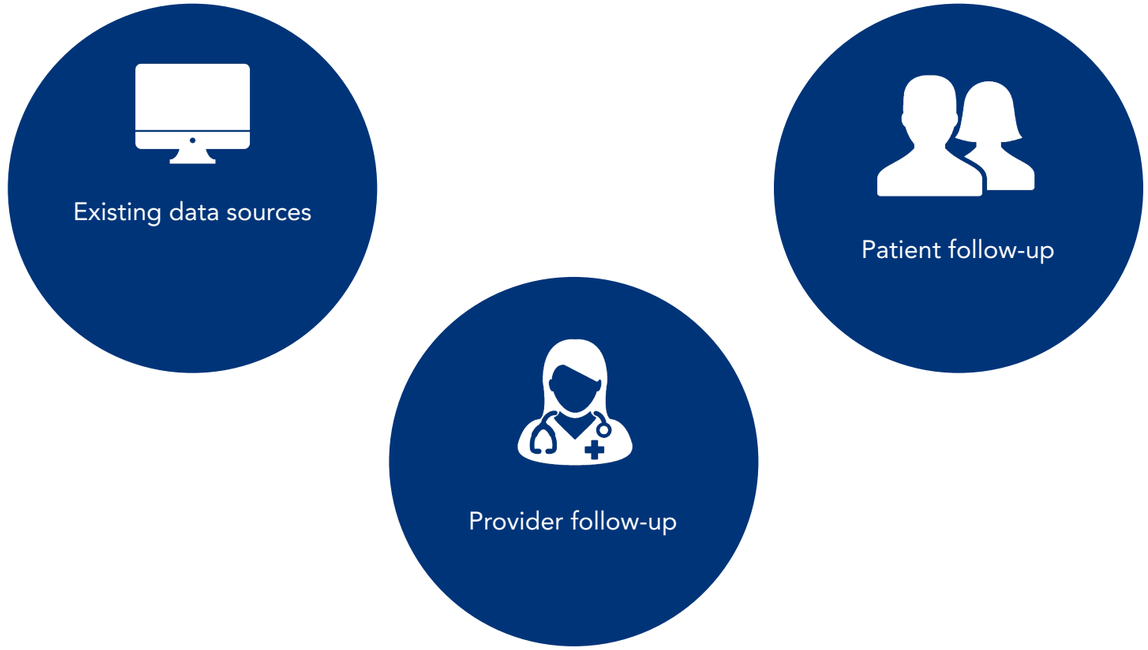
- ☑ Number (as well as gender) of all sex partners in the past 30 days
- ☑ Most recent sex partner history
  - Last sexual encounter
  - Race/ethnicity
  - Gender
  - Age
  - HIV status
  - Sites of exposure (Receptive and/or insertive; Vaginal, anal and/or oral)

Recommendation

If your project area already collects an "optional" variable, or an additional variable that is not included in the optional list but is already being collected (i.e., current use of PrEP, reason for testing, condom use, injection drug use, etc.), you are encouraged to continue collecting this information. Consider referencing the enhanced case-based population surveillance objectives.

How should data be collected?

Data collection will be different for every project area. Some common ways of collecting data are through:



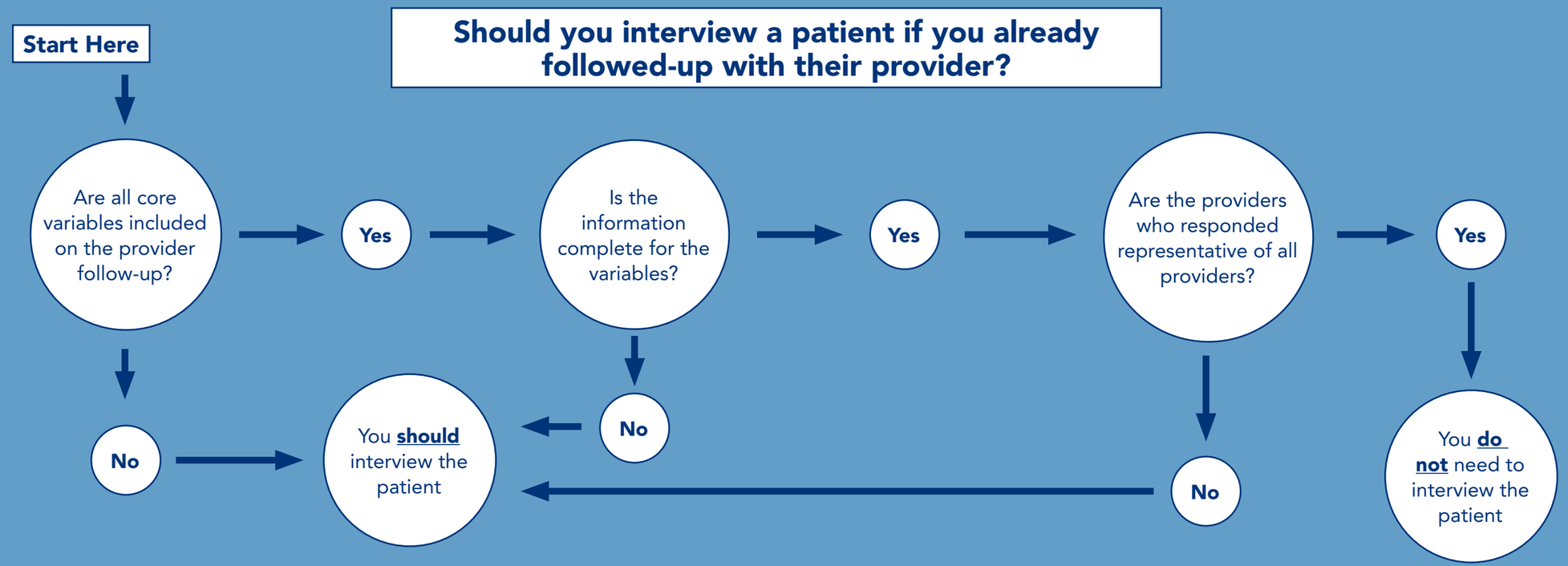
Recommendation

When possible, using existing data sources (i.e., sources for which a project area already has access) will save time and resources.

What are existing data sources for reportable GC cases?

Examples of existing data sources for reportable GC cases, or "internal case investigation", include data from STI clinic visit records, client medical records and "electronic charts", diagnosis records, laboratory records, treatment records, and facility reference files. These sources will provide you with initial case information as well as help you determine if a patient has previously been reported to the department of health and whether the record represents a 'duplicate record.' Project areas should report a positive GC laboratory report as a new case unless a prior GC infection has already been reported as a case and 1) the previously reported infection had a specimen collection date in the prior 30 days or had a documented treatment date in the prior 30 days and 2) there is no evidence of re-infection.

SECTION 4: COLLECTING DATA



What does *provider* follow-up entail?

Provider follow-up includes contacting the diagnosing provider to receive additional information about the case's clinical characteristics, the specific care setting and patient demographics that are not included in the original case or laboratory report. Common follow-up contact methods include phone calls, mail, secure faxes and secure emails that ensure patient information is protected. Those conducting provider follow-up may also use this opportunity to collect contact information for conducting patient follow-up, if this information is needed and missing from initial laboratory or case reports.

Project areas should do their best in determining the representativeness of the providers in which follow-up was conducted. If you are unable to determine if provider type is truly representative, be sure to note this when analyzing and interpreting your data.

What does *patient* follow-up entail?

Patient follow-up or interviews may be conducted by phone or in-person. Those conducting patient follow-up should make at least three documented attempts to contact each patient referred for follow-up. Sites must develop local protocol documents and data collection instruments for those conducting patient follow-up that ensure adequate training and address local human subject's requirements. If patient contact information is not included in existing sources, it may be found using vital record searches, registry searches, provider contact, social media (following local conventions), driver's license and/or vehicle registration registries.

Reminder
The same effort should be made to track down every case. For instance, if you decide you will call a patient six times, do that for every single case. If you treat cases differently, you will be introducing ascertainment bias, which may influence your data.

SECTION 4: COLLECTING DATA

From where should the data be collected?

Data sources will be different in every project area. Project areas are encouraged to use existing data first and then use additional data sources. Potential sources of data for the core variables listed above may include the following.

Data sources for specific core variables may include those listed in this table:

Potential Data Sources

	Laboratory data, provider and/or patient report	Laboratory data and provider report	Patient and/or provider report
Core Variables	☑ Age (date of birth) ☑ Sex ☑ Race/ethnicity ☑ County ☑ Gender identity/sexual orientation ☑ HIV status ☑ Previous history of GC infection ☑ History of substance abuse (IVDU, etc.) ☑ Date of treatment ☑ Co-infection with other STIs	☑ Specimen collection date ☑ Date of diagnosis **Note:** *if the person had a second infection after 30 days of an infection, they would be considered a new case* ☑ All anatomic site(s) of infection ☑ Gonorrhea-related sequelae (PID, disseminated gonococcal infection) ☑ Treatment provided (name and dose of treatment antibiotic) ☑ Diagnosing facility type (STI clinic, correctional facility)	☑ Sex of sex partner(s) ☑ Pregnancy status ☑ Clinical Symptoms and signs (health-care seeking behaviors; length of time symptoms were present) ☑ Partner treatment (EPT provision)

Video: Collecting Data for Enhanced GC Surveillance: Example from NYC

https://vimeo.com/372660503

SECTION 4: COLLECTING DATA

Video: Collecting Data for Enhanced GC Surveillance: Example from Philadelphia

https://vimeo.com/374513672

SECTION 4: COLLECTING DATA

AFTER DATA COLLECTION

How do I check for data quality?

- Align variable names and formats with any necessary requirements, if reporting to external organization
- Check that key identifiers align across datasets
- Address duplicate records
- Perform routine analysis of data to ensure representativeness
- Address missing data/ ensure complete data for core variables available on laboratory reports

Reminder

Increasing the percent completion of one or two variables (e.g., HIV status, pregnancy status) for all cases is NOT considered enhanced surveillance. Enhanced surveillance requires conducting provider follow-up and, if needed, brief patient interviews with a random sample of cases that results in complete information for all variables of interest for sampled cases.

TOOLS

- De-Duplication Guidance for Gonorrhea and Chlamydia Laboratory Reports
- Enhanced STD Surveillance Network (eSSuN) Protocol and Project Implementation Guide
- Example: Florida Interview Completion Tracking Sheet
- Maryland data collection instructions for enhanced GC surveillance
- Maryland data collection form for enhanced GC surveillance
- SSuN Cycle 4 Protocol Including Data Dictionaries
- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology

SECTION 5

Data Systems

How do I use existing data systems to conduct enhanced GC surveillance?

What data system/s does my site already use?

There is no need to reinvent the wheel. If you have a data system that works well for conducting other types of surveillance, start with investigating its use for enhanced surveillance!

How do I integrate enhanced GC surveillance systems with other types of surveillance systems (e.g., HIV or other STIs)

Considerations for integrating enhanced GC surveillance include:

1. Preventing duplication of work
2. Overlapping data needs
3. Overlapping partnerships

What are the unique considerations or gaps in existing data system/s in regards to enhanced GC surveillance?

- Data management systems, both vendor-based and locally built, can often be modified to randomly flag cases as there are entered into the system, which will make random sampling easier to conduct on an ongoing basis
- Automating processes like this supports prompt referral of cases for enhanced surveillance and maximizes the likelihood of successful follow-up
- Some project areas may face challenges modifying data systems to include additional and necessary variables.

If this relates to you, note that…

1. Many of the core variables are already included in the generic and STD Message Mapping Guides (MMGs) and the NETSS record layout version 5.0.
2. You may need to create a separate, locally maintained database. Be sure to follow your site's proper data security protocols for developing a new database. Database format options could include…
   - Access
   - Excel spreadsheets
   - Paper-based forms
3. Various user groups (e.g., MAVEN, PRISM, and NBS) are there to support you. The National Coalition of STD Directors (NCSD) holds user group calls for those using various data systems to discuss site needs and experiences. Contact Robin Hennessy at btj2@cdc.gov or Charlie Rabins at crabins@ncsddc.org to be added to one of those user groups.

Reminder

Reporting all core variables for enhanced surveillance is required for all project areas as part of STD PCHD surveillance component. Make sure to reach out to your prevention specialist if you have questions.

TOOLS

- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology

SECTION 6

Weighting Data

How do I make sure the data from my sample are representative?

What does "weighting data" mean?

When you obtain a random sample, you are getting a subset of all diagnosed, reported GC cases that met your inclusion criteria and were selected for interviews. This sample should reflect the overall population from which the sample was obtained in respect to key factors (e.g., age, gender, race/ethnicity). However, it is unlikely that all GC cases that were selected for interviews will respond, causing cases from the sample to be missing. Since you will only be able to analyze the cases in your sample that responded (completed interviews), you will need a way to account for these missing cases. Weighting data is a way to make sure the sampled cases align and are representative of the total population of reported cases when there are cases in the sample that are missing.

Why is it important?

The sample with which we are drawing inferences from will likely be a subset of cases from your original random sample, since it will only include those who completed the interviews. Using weights helps provide better estimates of the proportion and number of cases in the surveillance population by accounting for biases that may result from missing data.

When do you need to do it?

As discussed in Section 3: Random Sampling project areas must make sure that the sampled cases that are selected for enhanced GC surveillance are representative of the overall surveillance population by comparing the distribution of key demographics between sampled cases and the overall surveillance population.

However, it is likely that you will not be able to complete enhanced GC surveillance for all of the cases selected in your sample. Reasons for this include being unable to contact a patient for an interview or not receiving a provider response to multiple requests for complete clinical information. Whatever the cause, the sampled cases with completed investigations will become the sample you will use for data analysis.

If you are unable to complete enhanced GC surveillance for all of your sampled cases, weighting data makes it so the cases that do have completed data are representative of the total population of reported GC cases.

KEY TERMS

Selection (or inclusion) probability:
The probability that a case is included in the sample

Effective sample:
Sampled cases who completed interviews

Effective sample fraction:
The proportion of responding cases who completed interviews out of all reported GC cases in the project area

Design weight:
The number of all reported GC cases in the project area divided by the number of completed cases. Weights are interpreted as the number of cases each sample case represents.

Nonresponse weight:
Adjusts for nonresponse

How do you do it?

Overview of steps for weighting data

STEP 1: Check for nonresponse bias

Nonresponse bias occurs when people with certain characteristics, respond at a greater proportion than others. For example, if many more females compared with males respond to requests for interviews, this will lead to more missing data among males in your effective sample, which biases your sample.

To check for nonresponse bias, you will need to stratify the cases in your effective sample by variables of interest to see if certain types of individuals responded more or less frequently than others. If there is a meaningful difference in response by certain characteristics, then you will need to decide if you should adjust your case weights to account for these differences. As a general rule, a difference of 2.5% is often indicative that there may be a need to adjust your data. Please refer to the "Why Weight?" presentation in the Tools section for step by step instructions for how to weight a sample with nonresponse bias.

Reminder

If you randomly selected different volumes of cases across different geographic areas, you will need to apply weights to each region separately instead of applying the same weight to all cases. This will ensure that cases from different regions are counted proportionally to the volume of cases they represent.

STEP 2: Identify your case weights

Identifying case weights includes dividing the total number of reported GC cases by the number of GC cases for which you were able to complete enhanced GC surveillance.

Rules to note:

- Case weights are never zero (this would mean a group should not have been included in your surveillance population)
- Case weights are always positive
- Case weights can be fractions

STEP 3: Apply your case weights

Using case weights during analysis of completed case investigations will allow you to generate estimates of the proportion and number of cases in the overall population by characteristics of interest.

Different data management systems or statistical packages use different functions to apply case weights. See the instructional video below for an example of how to investigate if you need to account for non-response bias, and how to create and apply design weights.

EXAMPLE

Let's say there are 2,000 reported GC cases in your project area. You sampled 10%, so your random sample includes 200 cases. However, you were only able to obtain complete investigations for 160 cases.

$$\frac{\text{2,000 Reported GC Cases}}{\text{160 Complete Cases}} = 12.5$$

Your design weight is 12.5!

This means that each case will count 12.5 times when you analyze your data.

Reminder

For analysis, if you need to implement both a design weight AND a nonresponse weight for your cases, then the weights must be combined prior to applying them. Please refer to the "Why Weight" presentation in the Tools section for more guidance about combining and applying weights.

Video: Weighting Data for Enhanced GC Surveillance: Recommendations from CDC

https://vimeo.com/377346332

SECTION 6: WEIGHTING DATA

TOOLS

- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology
- Why Weight? Sample Weighting and Its Application to SSuN

R

- Design Weights
- Non-response Weights
- Survey Weights in R

SPSS

- SPSS Tutorials: Weighting Cases
- Weighting cases

Stata

- Choosing the Correct Weight Syntax
- Probability weights, analytic weights, and summary statistics

SECTION 7

Analyzing and Interpreting Data

What does gonorrhea look like in my community?

BEFORE DATA ANALYSIS

Develop a data dictionary

- This is especially important if someone else will be analyzing your data to make sure that the variables are understandable and the data will be analyzed and interpreted correctly
- This is also important if new variables are added to your data system and someone else is responsible for adding the new variables into the system

Clean your data

- Identify incorrect, incomplete, improbable and/or duplicate data
- Correct the incorrect, incomplete, improbable and/or duplicate data, if possible

Make sure the cases you wish to analyze are representative of your surveillance population

If not representative:

- Apply necessary case weights
- Apply other necessary methods to minimize bias

Recommendation
Consider using graphs, charts and descriptive statistics to summarize your data and make it easier to identify outliers.

DURING DATA ANALYSIS

Data should be monitored and tracked on an on-going basis. Sites should review their data annually at a minimum, but may wish to review their data more frequently, depending on available resources. This will help the areas better understand their communities and, at a broader level, support CDC with GC surveillance efforts.

Common and suggested calculations include…

- Total number of reported GC cases during a pre-defined time frame
- Total number of completed investigations, including completion rate
- Variable/data element completeness among investigated cases
- Summary/distribution of cases by core/ required variables, including the proportion of GC cases…

Reminder

Weighted data are analyzed differently than unweighted data. Incorporating weights into analyses will provide you with confidence intervals, which you will not get when you are analyzing unweighted data.

Different statistical programs have different procedures for weighted data. Feel free to review the "Tools" below for resources to assist you with analyzing weighted data.

REVIEW

What is stratification?

Stratifying data means splitting your data into groups so that you can look at different patterns among those with various characteristics. For example, you could look at the proportion of GC cases who identify as male, female and another gender separately, thus stratifying GC cases by gender.

Identifying key sub-groups in your population to assess

Conduct descriptive analyses to identify trends in GC in your project area, including…

- Counts
- Proportions
- Graphs

Stratify GC cases by key demographic variables including gender, race/ethnicity, age categories, and geographic region to see how proportions of GC cases differ by these variables.

SECTION 7: ANALYZING AND INTERPRETING DATA

What are the data telling me?

Measure of association: A measure of association allows you to compare disease occurrence across groups. There are many types of measures of association including rate ratios, risk ratios, prevalence ratios and odds ratios. The measure of association that should be calculated depends on the analytic study design you select. For enhanced GC surveillance, you will commonly use rate ratios to analyze your data.

Reminders

1. Association is not the same as causation. For instance, in the example above, it would be incorrect to say that being a man causes or leads to GC infection.
2. Associations may be confounded. **Confounding** occurs when a measure of association is influenced by a third variable that is associated with both the exposure and the outcome of interest.

EXAMPLE: RATE RATIO

If the rate of reported GC among:

- Men is 120 per 100,000 males
- Women is 80 per 100,000 females

$$\frac{120}{80} = 1.5$$

Your rate ratio is 1.5!

Interpretation: The rate of reported GC among men is 1.5 times the rate reported among women.

EXAMPLE: CONFOUNDING

The association between county (exposure) and reported GC rates (outcome) may be distorted by age (confounder) as GC is most common among young people and age distribution may differ by county.

For instance, let's say:

You want to know: Is the rate of reported GC higher in County A or County B?

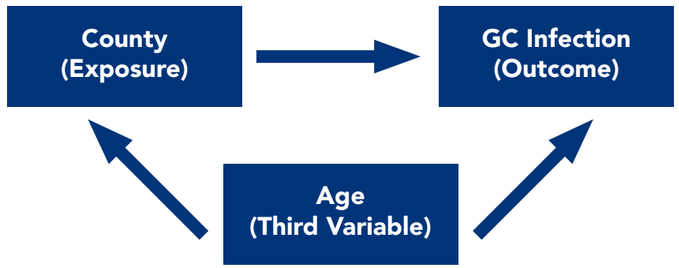
You see that: County A's overall reported GC is 2.5 times the rate in County B (Rate ratio = 2.5)

BUT…

County A's population is really young

County B's population is really old

Age is associated with both county and GC infection



If you stratified by age, rates of GC among younger people aged 15-24 may actually be similar in County A and County B. Age may be *confounding* the association between county and overall GC rates.

SECTION 7: ANALYZING AND INTERPRETING DATA

AFTER DATA ANALYSIS

How can the data be used for decision-making?

Consider what the data are telling you and how you can use this information to make decisions to reduce the spread of GC and improve health outcomes in your project area. Data driven reviews can be used to drive planning and implementation. Review Technical Assistance Note #17a | Data-driven reviews for more information.

TOOLS

- Data cleaning manual CDC
- Enhanced STD Surveillance Network (eSSuN) Protocol and Project Implementation Guide
- Example Table Shell for Enhanced Gonorrhea Surveillance
- Maryland Summary Report and Analysis
- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology
- Technical Assistance Note #16 | Epidemiologic analysis

Data for Decision Making

- Technical Assistance Note #17a | Data-driven reviews

R

- Design Weights
- Non-response Weights
- Survey Weights in R

SAS

- Maryland example SAS code for summary of cases

SPSS

- SPSS Tutorials: Weighting Cases
- Weighting cases

Stata

- Applied Survey Data Analysis in STATA 13
- Probability weights, analytic weights, and summary statistics
- Survey Data Analysis in STATA

SECTION 8

Disseminating Data to Improve Health Programs

How do I use knowledge gained through enhanced surveillance to improve programs?

What information should be highlighted to support project areas in making program improvements?

Think about whether there are…

Recommendation
Consider whether these align with gaps in screening or access to STI services.

How frequently should data be disseminated?

Analysis and dissemination are an ongoing process. To keep the process moving, you may wish to…

Hold routine meetings between data management, epidemiologic, and programmatic staff to:

- Review data
- Identify new analyses needed to inform decision-making
- Discuss factors that may be contributing to findings

"Routine" could mean quarterly or another timeframe that makes sense to your team.

Create an annual (or more frequent!) summary report with the most current data on GC in your project area and trends over time.

Share the report, along with any other data summaries you choose to develop, with key internal and external stakeholders.

Remember, if you excluded some cases that were not eligible for enhanced surveillance (e.g., incarcerated populations), be sure to note this when disseminating your findings (i.e., add a footnote to tables and figures).

Work with partners to understand the frequency of dissemination that is most helpful for them

Who are the key internal and external stakeholders for your project area?

Internal stakeholders may include:

- Staff in the STI program (such as DIS)
- Other health department sections (such as reproductive health or school health)
- Health department leadership

External stakeholders may include:

- Medical providers
- Laboratories (public and private)
- Policymakers
- School officials

Reminder

Frequent and timely dissemination helps stakeholders understand disease trends, transmission, and outbreaks, which helps them respond effectively.

How should you share the information you learn through enhanced GC surveillance with stakeholders?

There are federal, state, and local policies for data sharing to ensure security and confidentiality.

Work with stakeholders to learn their data sharing protocols and the formats that are most useful to them. Some helpful tips include:

- When it comes to tables and figures, simplicity is key
- Include summarization and/or interpretation with tables and figures to make sure they are not misinterpreted
- Include slide decks when disseminating summary reports if this is helpful to stakeholders

Reminder

What are your project area's federal, state, and local policies for data sharing? Make sure to follow them!

- Consider the types of materials that will be most useful for specific audiences and tailor your materials to fit their needs
  - i.e.: Will healthcare providers read a full report or would they prefer a fact sheet with just the key information they need to know to support their patients?
- When in doubt, ask stakeholders what kinds of data and in what format are most useful to them
- Think about new and creative ways to communicate information, such as infographics or fact sheets
- Consider presenting your findings at local and national scientific conferences and/or publishing your findings in peer-reviewed journals and reports
- **Extra credit:** Develop an online query system for stakeholders to access data

Reminder

Stakeholder engagment

Consider how stakeholders should be engaged throughout the data dissemination process. They will provide crucial insight about the information they need, how often they need, and what format would be most helpful for them!

Your time is valuable. Do not waste it by creating complicated data reports that are run frequently and never reviewed. Make sure the dissemination materials you are spending time developing meet the needs of key stakeholders so that they are useful for decision-making.

TOOLS

- California STD Infographic
- California STD Annual Surveillance Report Slides
- California 2017 Snapshot: Syphilis and Gonorrhea among Men who have Sex with Men (MSM)
- Maryland Summary Report and Analysis
- Michigan 2018 STD Fact Sheet
- Michigan 2018 STD Overview Slide Deck
- SSuN Project Summary Report
- SSuN Washington Population Report
- Technical Assistance Note #16 | Epidemiologic analysis

SECTION 9

Implementing Enhanced GC Surveillance

What does it take to start enhanced GC surveillance?

What staffing capacity is needed?

What is a reasonable number of full-time employees (FTEs) for conducting enhanced GC surveillance?

What are current FTEs able to take on?

Consider

- Your project area's needs (different for low vs. high morbidity project areas)
- Your staffing needs
  - Are your current FTEs available to provide support?
  - Will you need to hire additional FTEs, contractors or student interns to take on the extra work load?
  - Should your enhanced GC surveillance period be adapted to account for staff workload?
- Your available resources

What skills or qualifications are necessary for staff to have?

- Proficiency in Microsoft Excel and/or statistical software (e.g., SAS, R, Stata, SPSS)
- Ability to conduct basic data analysis
- Demonstrated experience managing and conducting data collection activities, including interviewing experience
- Strong written and verbal communications skills
- Strong organizational skills

Recommendation

Think about whether you are able to budget for training and workforce development for existing staff. Building skills needed for enhanced GC surveillance, including use of statistical analysis software, may benefit other program areas as well!

Examples of Enhanced GC surveillance across varying morbidity levels

Low	Middle	High
412 cases	10,000 cases	30,000 cases
DIS conduct interviews	Contractors conduct interviews (e.g. graduate students/fellows)	Epis, contractors, and DIS conduct interviews
Interviewed all cases	Interviewed 30% of cases	Interviewed 10% of cases
Entire state included	5 counties included	1 county included

**Table provides representative examples; methodology should be tailored to each project area's available resources and organizational structure*

What should implementation of enhanced GC surveillance look like?

Project areas are encouraged to take a phased implementation approach, where your site's annual strategy will depend on its level of experience performing enhanced surveillance. It may take several years before enhanced surveillance is fully implemented.

For instance, if your project area has never performed enhanced surveillance before you will likely have to start by getting your infrastructure in place, such as making modifications to data systems so that you are able to select a representative random sample from your surveillance population, developing protocols for assigning selected cases to staff for investigation, deciding how data will be collected and input into data systems, piloting all data collection tools, determining which stakeholders are important for your project area, and assessment of staffing needs. It is likely that you will also want to develop a timeline for implementation so that you know what resources you need for each implementation phase.

Reminder

Randomization

Remember that you must develop your methods for selecting your random sample of GC cases prior to implementing enhanced GC surveillance.

Example: Enhanced GC Surveillance Implementation Plan for Project Area with Limited Experience

Year 1	Year 2	Years 3 – 5
Step 1: Assess morbidity & determine geographic area	Step 1: Gain local support	Step 1: Conduct investigations
Step 2: Determine appropriate timeframe	Step 2: Finalize data collection tools identify/train staff	Step 2: Conduct non-response analysis to monitor implementation
Step 3: Develop and validate methods for taking a random sample of GC cases	Step 3: Modify surveillance information systems to store data	Step 3: Analyze data to generate representative estimates
	Step 4: Pilot investigations	Step 4: Ongoing implementation & quality assurance

TOOLS

- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology
- Technical Assistance Note # 2 | Enhanced surveillance for GC cases

SECTION 9: IMPLEMENTATION

SECTION 10

Putting it all Together

Remember that enhanced GC surveillance will look different for every project area and may change over time.

Some common challenges and key considerations among sites include…

QUESTION **How long will it take to implement enhanced GC surveillance?**

ANSWER Project areas should take a phased approach to implemented enhanced GC surveillance. How quickly implementation happens depends on the project area's existing experience and resources.

QUESTION **What if it is not sustainable to conduct enhanced GC surveillance throughout the whole year due to personnel capacity?**

ANSWER You are not expected to conduct enhanced GC surveillance throughout the whole year if that is not feasible for your project area. However, it is important to select a time period that will produce reliable data. Any period shorter than three months may not be reliable and is not recommended.

QUESTION **How do I transmit data from enhanced GC surveillance to CDC?**

ANSWER Although there will be differences across STD PCHD sites, note that…

1. Many of the variables outlined in Section 4: Collecting Data for Enhanced Gonorrhea Surveillance are included in the generic and STD Message Mapping Guides (MMGs) and the NETSS record layout version 5.0
   - These resources have a variable (ie: CASESAMPLE) that indicates if a case was randomly sampled for enhanced investigation — be sure to mark "yes" for all randomly sampled cases, even if patient interviews were not completed → this will let CDC colleagues have a sense of which cases were selected for enhanced surveillance
2. All cases that were randomly selected for enhanced GC surveillance should have all information that was collected during enhanced surveillance added to their case records — this information should be reported to CDC
3. Additional information may be shared through various formats including:
   - Progress reports (APRs, IPRs)
   - Workplan updates
   - Targeted Evaluation Plans (TEPs)

Final reminders

You are expected to...

- Identify a geographic area targeted for enhanced surveillance in your project areas and capture data for all core variables for all GC cases (or a representative sample of all GC cases) in that area
- Use provider follow-up, patient interviews, and other methods you have access to in your project area to capture these data
- Conduct enhanced surveillance in this geographic area for a specific period of time (e.g., six months)

You are not expected to...

- Follow up on all GC cases in all areas in your project area
- Conduct enhanced GC surveillance throughout the whole year if that is not feasible for your project area

Reminder

Stakeholder engagement

Remember to engage stakeholders as much as possible throughout enhanced GC surveillance activities. Different stakeholders will be able to provide insight regarding:

- best ways to reach individuals to collect data
- data interpretation
- dissemination approaches

TOOLS

- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- SSuN Best Practice Note: Data Management

Tool Collection

Enhanced surveillance technical assistance resources

- STD PCHD 19-1901: Enhanced Surveillance for Gonorrhea Webinar
- Technical Assistance Note # 1| Case-based surveillance of STDs (CT, GC, syphilis)
- Technical Assistance Note # 2 | Enhanced surveillance for GC cases
- Technical Note # 2b | Enhanced surveillance of GC cases: Methodology
- Technical Assistance Note #16 | Epidemiologic analysis
- Technical Assistance Note #17a | Data-driven reviews

Surveillance resources

- 2017 STD Surveillance Report:
- NCSD User Groups (Maven, PRISM, NBS): Contact Robin Hennessy at btj2@cdc.gov or Charlie Rabins at crabins@ncsddc.org to be added to one of these user groups.
- STD Surveillance Coordinator's quarterly calls: Contact Ashley Vineyard at avineyard@cste.org to be added to the list
- The STD Surveillance Network (SSuN) resources:
  - Enhanced STD Surveillance Network (eSSuN) Protocol and Project Implementation Guide
  - SSuN Best Practice Note – Strategy B: Random Sampling

Data collection

- Maryland data collection instructions for enhanced GC surveillance
- Maryland data collection form for enhanced GC surveillance

Data cleaning, weighting and analysis resources

- Data cleaning manual CDC
- Example Table Shell for Enhanced Gonorrhea Surveillance
- Maryland Summary Report and Analysis
- SSuN Cycle 4 Protocol Including Data Dictionaries
- Why Weight? Sample Weighting and Its Application to SSuN

R

- Design Weights
- Non-response Weights
- Survey Weights in R

SAS

- Maryland example SAS code for case selection
- Maryland example SAS code for summary of cases
- Michigan example SAS code for sampling

SPSS

- SPSS Tutorials: Weighting Cases
- Weighting cases

Stata

- Applied Survey Data Analysis in STATA 13
- Choosing the Correct Weight Syntax
- Probability weights, analytic weights, and summary statistics
- Survey Data Analysis in STATA

Data dissemination and visualization resources

- California STD Infographic
- California STD Annual Surveillance Report Slides
- California 2017 Snapshot: Syphilis and Gonorrhea among Men who have Sex with Men (MSM)
- CDC STD Prevention Infographics
- Michigan 2018 STD Fact Sheet
- Michigan 2018 STD Overview Slide Deck
- SSuN Project Summary Report
- SSuN Washington Population Report
- Stephanie Evergreen Data Visualization Checklist

Resources from project areas

- Example: Florida Interview Completion Tracking Sheet

Data management and submission

- SSuN Best Practice Note: Data Management

Other resources

- De-Duplication Guidance for Gonorrhea and Chlamydia Laboratory Reports
- Gonorrhea — CDC Fact Sheet
- NCHHSTP Atlas Plus
- Principles of Epidemiology in Public Health Practice, Third Edition. An Introduction to Applied Epidemiology and Biostatistics
- Sources for data on social determinants of health

CONTACT

To inquire about implementation of enhanced GC surveillance in STD PCHD, please reach out to your project areas CDC prevention specialist

To inquire about referrals to project area in SSuN that may be able to provide technical assistance with enhanced GC surveillance:

Mark Stenger, CDC Project Officer
Email: zpl4@cdc.gov
Phone: 404-639-6136