

## Collecting Client-level Data

Client-level data collection involves gathering data about each individual client and maintaining that information in a database. Client data can then be retrieved, sorted, grouped, and analyzed across different variables of interest. In contrast, aggregate data collection combines information about all clients served by an intervention and does not retain client-specific data in a database. Client-level data can be pooled to yield aggregate data; however, information collected in aggregate form cannot be converted to client-level data.

Health departments can decide if they want to collect client-level data and for which interventions. Client-level data collection is usually limited to group level interventions (GLI), individual level interventions (ILI), and prevention case management (PCM) because these interventions usually provide sufficient interaction with clients to collect this information. Client-level data is rarely, if ever, collected during outreach or health communication/public information (HC/PI) interventions.

Client-level data collection typically involves assigning a unique identifier or code to each client. Linking a client's code and the client's data permits tracking of the individual client over time, as well as the aggregation, analysis, and reporting of data from multiple clients. The client code may be included on questionnaires and other forms for collecting data on client demographics, risk behaviors, intervention services received, and other variables of interest. Client codes can be generated by the client or the contractor.

Client-Generated Codes: The client creates a code by responding to a series of prompts, such as client initials, birth date, and mother's first name. With client-generated codes, the code is designed so that clients know all the information needed to complete the code themselves, though contractors may assist if necessary.

Contractor-Generated Codes: The contractor assigns a code to a client based on a series of prompts, such as provider initials and a number for each consecutive client seen by the provider. With contractor-generated codes, the contractor must create the code for the client because the client may not have all the information needed (e.g., provider initials). A master list is often maintained linking client codes and client names to ensure that clients are assigned the same code during subsequent contacts.

## Client Code Examples

Methods used by different jurisdictions to create client codes are described below. Following these, a method suggested by the Health Resources Service Administration (HRSA) is described. The examples provided are for a white male, non-Hispanic client named John Doe, born on March 16, 1963. John is the fifteenth client served by a provider named Mary Smith.

<b>Examples of How to Create Client Codes</b>			
<b>Jurisdiction</b>	<b>Who Creates Code</b>	<b>How Code is Created</b>	<b>Example</b>
Virginia	Client	1st and 3rd letter of first name, 1st and 3rd letter of last name	JHDE
Maryland	Client	birth month, birth day, complete birth year, gender, race, ethnicity	03161963MWN
New Jersey	Client	1st and 3rd letter of first name, 1st and 3rd letter of last name, birthmonth, last two digits of birth year	JHDE0363
Wisconsin	Contractor	Provider initials, consecutivenumber from the first client	MS015

HRSA creates a client code, called a Unique Record Number (URN), using the following method: 1st and 3rd letter of first name (if blank, use the middle initial), 1st and 3rd letter of last name (if blank, use the middle initial), birth month, last two digits of birth year, and gender code (1=male, 2=female). For example, JHDE03631. After this number is created, it is encrypted, or scrambled, using a complex algorithm. The resulting nine-digit code does not resemble the original information in any way. It is virtually impossible to retrace the information in the URN or retrace any personal information about a client. Decoding a URN is not feasible; too much of the original information is removed during the encryption process to be able to work backwards to the original 11 digit information.

### **Client Confidentiality**

Concerns about confidentiality can hinder efforts to collect client-level data and should be considered. Client codes typically avoid using complete names, portions of social security numbers, or any other information that may reveal the client's identity. Even in the absence of information that could reveal client identity, clients may perceive the potential for breeches of confidentiality and therefore be hesitant to report risk behaviors or to utilize prevention services that collect client-level data. These concerns may be particularly salient for clients engaged in illegal or stigmatized behaviors. Contractors may also be concerned about confidentiality issues and resist collecting client-level data.

Confidentiality concerns can be addressed in different ways. One health department conducted focus groups with clients and learned that they would feel more comfortable if contractors did not see client-level data. In this jurisdiction, clients generate their own code and complete questionnaires. These questionnaires are placed in a sealed envelope, which the contractor collects without seeing the information and sends to the health department for data entry and analysis. In a different jurisdiction, clients did not want the health department to have access to client-level data. In this case, the contractor collects and aggregates client-level data. Only aggregate reports are

submitted to the health department. Both approaches show a positive response to the particular concerns in each jurisdiction.

### Benefits of Client-Level Data

Collecting client-level data facilitates reporting several process monitoring data elements required by the Guidance, including client race, ethnicity, gender, and age; risk population served; number of clients served; and number of intervention sessions received for ILI, GLI, and PCM. ... In the absence of client-level data, health departments may not be able to report this information accurately. These data also may be useful for local evaluation and planning purposes beyond the Guidance requirements.

The advantages of client-level data are contrasted with aggregate data in the following example. A three session GLI targeting heterosexuals serves six clients. Jurisdiction A collects aggregate data, and Jurisdiction B collects client-level data. Both jurisdictions collect data on risk, race, gender, and the number of intervention sessions received.

Aggregate Data Collection: Jurisdiction A collected and reported the following aggregate data upon completing the three sessions GLI.

<u>Number of clients attending intervention sessions by risk, race and gender</u>			<u>Number of clients attending the first, second and third intervention session</u>		
<u>Risk</u>	<u>Race</u>	<u>Gender</u>	<u>First</u>	<u>Second</u>	<u>Third</u>
5 Hetero 1 MSM	3 White 3 Black	3 Male 3 Female	6	3	3

In this example, aggregate data do not permit reporting of client race by gender because these demographic data were collected and reported independently. There is no way to identify the race of the one MSM client served by the intervention nor is it possible to know how many intervention sessions were received by each client (i.e., only the number of clients attending each session is known). Without this information it is difficult for Jurisdiction A to report all required Guidance data.

Client-Level Data Collection: Jurisdiction B collected and reported the following client-level data upon completing the three session GLI.

<u>Client</u>	<u>Risk</u>	<u>Race</u>	<u>Gender</u>	<u>Number of sessions completed</u>
Client 1:	Hetero	White	Female	2
Client 2:	MSM	Black	Male	3
Client 3:	Hetero	Black	Male	3
Client 4:	Hetero	Black	Female	3
Client 5:	Hetero	White	Male	2
Client 6:	Hetero	White	Female	2

In contrast to aggregate data, client-level data reveals greater demographic detail about clients served. These data show that the one MSM client was Black and that none of

the White clients received more than two intervention sessions. These data also reveal that three clients received two intervention sessions and three clients received three intervention sessions. This more detailed information makes it easier for Jurisdiction B to report Guidance data.

Client-level data can reduce the need for contractors to collect data each time they see a repeat client. For example, if contractors collect demographic and other data during the first contact with a client, they can then use the client's code to access the information for reporting subsequent visits.

### **Challenges of Client-Level Data**

Client codes must protect client confidentiality. Ideally, these codes are a unique, unduplicated identifier for each client. Using last names or portions of social security numbers (e.g., last four digits) in the code can decrease the possibility of code duplication. However, these elements are usually avoided to mitigate client concerns about confidentiality. Efforts to maintain confidentiality can make it difficult to avoid duplication, reducing the quality of client-level data.

Codes may also be unstable over time for the same client. For example, a client may have multiple names or nicknames that undermine the consistency of a code that uses initials or letters from their name. It may not be possible, therefore, to create client codes that eliminate the possibility of duplication and that are completely stable over time. Duplicate and unstable codes can compromise data quality. Health departments should be mindful of these problems and try to minimize their occurrence.

**Source:** Chapter 5, *Evaluation Guidance Handbook: Strategies for Implementing the Evaluation Guidance for CDC-Funded HIV Prevention Programs*, National Centers for Disease Control and Prevention (CDC), March 2002