

STARTING THE PROCESS

-
- Section 1 Determining the Scope
 - Section 2 Determining the Content and Organization of the Profile
 - Section 3 Determining the Development Process
 - Section 4 Obtaining Profile Data
 - Section 5 Identifying Skills for Preparing Profiles
 - Section 6 Understanding Basic Analytic Concepts

Chapter 2 gets down to the nuts and bolts: How do you begin to develop an HIV/AIDS epidemiologic profile? How do you determine the scope, content, and organization of the profile? What skills are required to prepare it? What data do you include? Where can you find those data? Once you get the data, how should you analyze and interpret them? Where can you get help? How do you address differences between prevention and care guidelines and differences such as service area boundaries, time frames, and due dates?

This chapter provides guidance for answering these questions. Of particular importance are the basic concepts and recommended methods for analyzing the profile data.

Establishing the foundation, presented in this chapter, is critical for developing consistent profiles that meet the specific needs of prevention and care planning groups.

Note. In this document, the term *HIV/AIDS* is used to refer to three categories of cases: (1) new diagnoses of HIV infection only, (2) new diagnoses of HIV infection with later diagnoses of AIDS, and (3) concurrent diagnoses of HIV infection and AIDS.

Section 1: Determining the Scope

The first step in preparing an HIV/AIDS epidemiologic profile is to determine its scope. The scope should be broad enough to provide planning groups with the information and background data they need to identify and set priorities among HIV/AIDS prevention and care services. At the same time, the scope needs to be narrow enough to meet specific requirements of prevention and care programs.

The appropriate scope of the epidemiologic profile depends on several factors outlined in this section.

CDC and HRSA Considerations

Although the data required are similar, CDC and HRSA have different requirements for developing and updating the HIV/AIDS epidemiologic profile. Below are several suggestions:

- Every year, the health department should update the executive summary and core epidemiologic data—including tables and figures—to ensure that planning groups can identify and set priorities among populations and their prevention and care needs.
- Less often, the health department should comprehensively revise its epidemiologic profile. CDC prevention planning groups should complete such a revision at a minimum of once every 5 years. CARE Act planning consortia and councils should follow their yearly program guidance from the HIV/AIDS Bureau.

Consider including as much as possible of the sociodemographic and socioeconomic information covering the entire service area. Updated profiles will then need to include only the data from those areas in which significant changes have occurred.

Factors Affecting the Scope

Scope refers to the boundaries, such as the time frame and geographic area, which define the extent of information in your profile. Determining the scope of your epidemiologic profile is a **collaborative** effort that requires consultation **with your planning groups and other potential stakeholders**.

Because each jurisdiction's needs differ, it is not possible to say how much time or how many resources should be allotted to complete an epidemiologic profile. The scope of your profile will affect the time and resources needed to complete it. It will be important to factor the time and resources needed into the planning process.

The following is a checklist of questions to guide your data gathering and analysis. Your answers will enable you to determine

- the geographic boundaries of the area described in your profile
- the extent to which the profile can address the core epidemiologic questions
- any special considerations pertinent to your service area and planning group requirements
- the time and resources needed to complete the profile

Checklist for Determining the Scope of an Epidemiologic Profile

- ✓ Is the profile a comprehensive epidemiologic profile or an annual update?
- ✓ What planning jurisdiction(s) or service area(s) should be included?
- ✓ What specific questions for prevention and care planning should be addressed by this profile?
Examples
 - special needs of populations at risk of becoming infected with HIV
 - trends in the epidemic that can be identified through a comparison of HIV and AIDS prevalence data
 - setting priorities for prevention and care services among prioritized populations
- ✓ What resources—time, personnel, and funds—are available to develop the profile?
- ✓ What sources of information are needed to answer the profile's epidemiologic questions?
Example
 - If your service area does not have HIV reporting, you will need to use estimates of HIV prevalence derived from AIDS case reporting.
- ✓ What data are available to supplement the HIV/AIDS Reporting System data that describe the HIV-positive population?
- ✓ What information is already available? More information, which is readily available, may allow for a broader scope.

Section 2: Determining the Content and Organization of the Profile

To be useful for prevention and care planning, a full epidemiologic profile should answer several core epidemiologic questions:

- What are the sociodemographic characteristics of the general population in your service area?
- What is the scope of the HIV/AIDS epidemic in your service area?
- What are the indicators of risk for HIV infection and AIDS in the population covered by your service area?

It should also answer questions specific to prevention or care planning needs, such as

- What are the patterns of service utilization of HIV infected persons in your area?
- What are the number and characteristics of persons who know they are HIV-positive but who are not receiving primary HIV medical care?

The material should be contained in sections organized in a logical sequence that allows end users to locate information quickly. Chapter 3 explains how to answer the core questions and the questions specific to care programs. The remainder of this section describes the organizational framework of an effective and user-friendly profile.

Profile Sections and Organization

As is true of any good document, a well-organized profile is divided into logical sections:

- front matter
- introduction
- body
- conclusion
- appendixes
- other back matter

Front matter

The front matter should consist of the following:

- **Contributors**, a list that includes the names of writers and others who worked on the profile
- **Abbreviations**, a list of the shortened names for terms and organizations that appear in the profile
- **Executive summary**, a synopsis of the profile's content
- **Table of contents**, a listing of, and page numbers for, topics, tables, and figures

Introduction

The introduction should include the following:

- **Background** about the history and purpose of the profile
- **General description** of data sources and their strengths and limitations to ensure that users understand what the profile can and cannot explain
- **Overall description** of the profile’s strengths and limitations
- **Preparation information** that describes the development guidelines, process, key players, and how the development of the profile followed the guidelines and process

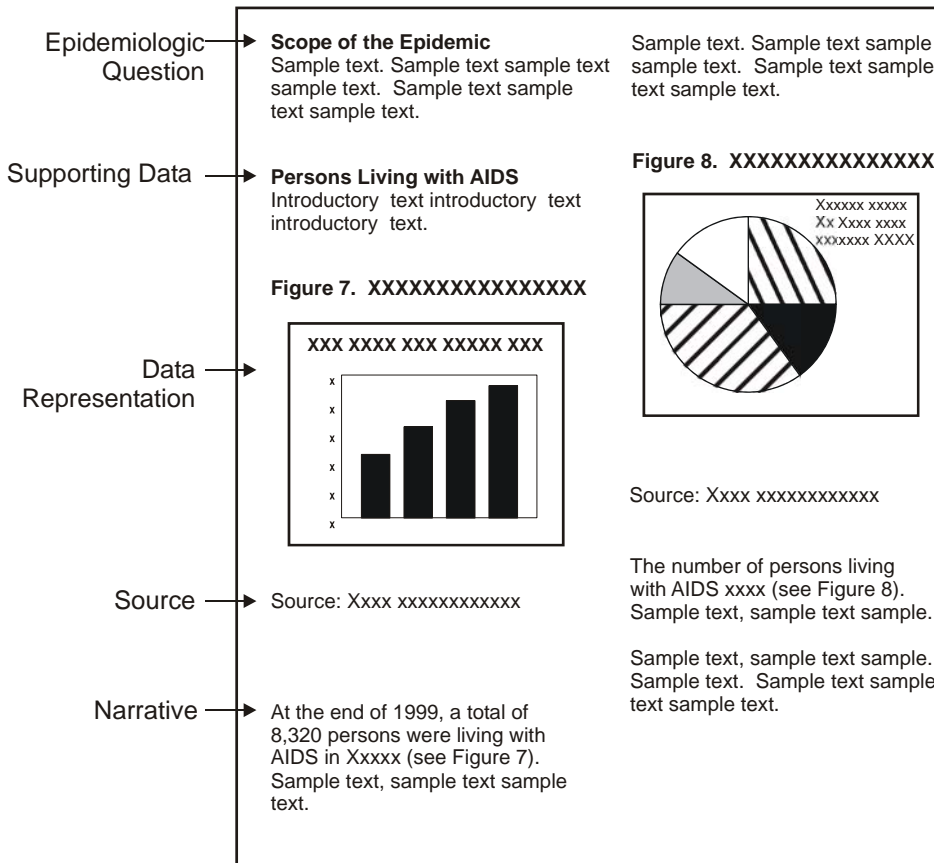
Body

The body of the profile includes the epidemiologic questions and the data that answer the questions.

Data are typically presented in tables, graphs, pie charts, or maps. These presentations should be accompanied by a narrative that explains and expands upon the data.

Figure 2-1 illustrates the content requirements of a page in the profile. No particular page layout for the profile is preferred. The one below is just an illustration.

Figure 2-1
Example of layout of profile body



Conclusion, appendixes, and other back matter

The back matter should consist of the following:

- The **Conclusion** summarizes the data and trends and highlights key findings.
- **Appendixes** contain information on data sources, supporting documentation, and a feedback form for end users to complete and return to the authors.
- **Other back matter (in addition to the appendixes)** includes items such as a glossary and a list of references or suggested readings.

Section 3: Determining the Development Process

This section outlines a recommended process for developing the epidemiologic profile. It presents a logical and ideal flow. You may find that the process you use to develop your profile is different and that you need to revisit some steps as you go along. The key is to incorporate the principles of the process to ensure that your profile is comprehensive and the presentation is of high quality.

Development Process

There is a 7-step process recommended for developing the epidemiologic profile (see box below). One of the keys to a comprehensive and user-friendly profile is to ensure that the development process is a collaboration between you, as the writer(s), and the planning group. This process may be different for planning groups that have merged.

At the beginning of the development process, it would be useful for the writer(s) to meet with the planning group to

- get to know the dynamics of the group and the challenges they face
- introduce the new guidelines and why HRSA and CDC developed them
- determine the planning group's experience with past profiles and where improvement is needed
- determine whether the group has special requests for data or interpretation

Recommended Development Process

1. In collaboration with state and local surveillance staff and prevention and care planning groups, determine the specific and unique needs of the planning group. Determine the boundaries and the scope of the profile. Establish mechanisms to ensure collaboration throughout the process.
2. Obtain core and supplemental data. Determine which of these data to include in the analysis.
3. Analyze and interpret data.
4. Present data in user-friendly formats.
5. Draw overall conclusions and write an effective, useful narrative.
6. Write remaining sections and compile complete epidemiologic profile.
7. Prepare clear presentations for appropriate audiences.

Here are some additional tips to help ensure a successful profile development process:

- realize that the process is a group effort; it cannot be done by one person
- expect that a full-time equivalent staff person will need at least several months to complete the profile
- have a knowledgeable person with technical expertise on local data sources review and proofread the document
- create a dissemination plan well before your profile is complete and ready for distribution

Section 4: Obtaining Profile Data

The next step in preparing the profile is to obtain the data you will use to address the scope of the profile and answer the epidemiologic questions.

As the profile writer, you need to be aware of several considerations concerning the acquisition and use of data to describe the epidemic in a service area. You also need to know what types of data are available and where to obtain them.

This section presents a *general* discussion of those considerations—the types of available data and where to find them—as preparation for Chapter 3, where these issues are applied to the specific core epidemiologic questions.

General Data Considerations

The following are considerations for reviewing data and data sources that you may use in the epidemiologic profile:

- **Completeness of the data:** How well does the number of reported HIV or AIDS cases reflect the true number of persons who have HIV infection or AIDS and are thus eligible to be reported? For example, how well does the prevalence of AIDS represent the true number of persons living with AIDS in your service area?
- **Representativeness of the data:** How well do the characteristics from a data source correspond to the characteristics of the overall population? For example, data from a hospital-based sample may not represent all HIV-infected persons or all HIV-infected persons in care in the area covered by the survey.
- **Age of the data:** How old are the data that will be used for analysis? For example, a behavioral survey conducted in 1990 might not provide data that are sufficiently up-to-date for current prevention activities.
- **Timeliness of the data:** How long is the reporting delay between the diagnosis of HIV or AIDS and the report to the health department?
- **Limitations of the data source or variable of interest:** Consider the limitations of the data source or variable. For example, AIDS case data are the only HIV-related data that are consistently available on a population-wide basis in all states by sex, race/ethnicity, age, and mode of HIV exposure. However, AIDS case data may not reflect the characteristics of people who were recently infected with HIV.

- **Surrogate, or proxy, markers:** A proxy variable is used as a marker for other variables when what we really want to measure is too difficult to measure directly. For example, some areas may use sexually transmitted disease (STD) data as a proxy when data on sexual behaviors are not available.
- **Validity of the data:** How well does a variable measure what it is intended to measure? For example, how well was information about age transcribed to the case report from the medical record (how accurate are the case report data compared with those in the medical record)?
- **Small numbers:** You may need technical assistance to interpret the data when analyzing small numbers of cases because small absolute changes in the number of cases can produce large relative or proportionate changes in rates that may be misinterpreted by end users. These analyses may also require the use of advanced statistical tests. Rates calculated from numerators smaller than 20 should be denoted in a footnote as unreliable.

Types and Sources of Data for Epidemiologic Profiles

This section includes a description of commonly available data and their sources. Several of these sources directly report HIV and AIDS cases and clinical conditions of persons with a diagnosis of HIV infection or AIDS. Other sources are used to round out the picture of the HIV/AIDS epidemic in your service area. Other sources also are used if no HIV incidence data are available. See Appendix A for an expanded list of core and supplemental data sources and references.

Type of Data	Description	Where to Obtain
AIDS surveillance	<p>AIDS reporting began in 1981, and AIDS is a reportable condition in all states and territories. The AIDS surveillance system was established to</p> <ul style="list-style-type: none"> • monitor incidence and the demographic profile of AIDS • describe the modes of HIV transmission among persons with AIDS • guide the development and implementation of public health intervention and prevention programs • assist in the evaluation of the efficacy of public health interventions <p>State and local health departments actively solicit disease reports from health care providers, laboratories, and other sources. Standardized case report forms are used to collect sociodemographic information, mode of exposure, testing history, and clinical information. AIDS surveillance has been determined to be more than 85% complete.</p>	<p>All 50 states, the District of Columbia, and US territories collect AIDS surveillance data. Contact your state or local service area's HIV/AIDS surveillance coordinator.</p>
HIV surveillance	<p>HIV surveillance data include all persons who meet the 1999 case definition for HIV infection and have been reported to a state or local health department. HIV</p>	<p>As of January 2004, 34 states (Alabama, Alaska, Arizona, Arkansas, Colorado, Florida, Georgia, Idaho, Indiana, Iowa,</p>

Type of Data	Description	Where to Obtain
	<p>surveillance data</p> <ul style="list-style-type: none"> • provide a minimum estimate of the number of persons with a diagnosis of HIV infection whose test was confidential • identify emerging patterns of transmission • help detect trends in HIV infections among populations of particular interest (e.g., children, adolescents, women) that may not be evident from AIDS surveillance data <p>HIV surveillance data also provide a basis for establishing and evaluating linkages to the provision of prevention and early intervention services. They can be used to anticipate unmet needs for HIV care. According to state evaluations, HIV infection reporting is estimated to be 80%–90% complete for persons who have tested positive for HIV.</p>	<p>Kansas, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Nebraska, Nevada, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, Wisconsin, Wyoming), American Samoa, Northern Mariana Islands, Puerto Rico, the Virgin Islands, and Guam have implemented HIV case surveillance, using the same confidential system for name-based case reporting for HIV infection and AIDS.</p> <p>Pennsylvania implemented name-based reporting in areas outside the city of Philadelphia. Connecticut implemented mandatory HIV reporting in January 2002. For adults and adolescents 13 years of age and older, reporting is by name or code (if patients or physicians prefer this method). For children < 13 years of age and for persons who are co-infected with tuberculosis, reporting is by name. New Hampshire allows HIV cases to be reported with or without a name. Five states use names to initiate case reports and then convert to a code (Delaware, Maine, Montana, Oregon, Washington), and 10 areas are using a coded identifier rather than patient name to report HIV cases (California, Hawaii, Illinois, Kentucky, Maryland, Massachusetts, Rhode Island, Vermont, and the District of Columbia).</p>
Behavioral surveillance	<p>Data on behaviors that are relevant to HIV prevention, transmission, and medical care are available from a variety of sources, including general population surveys, surveys of populations at risk for HIV, and surveys of persons with HIV or AIDS. Behavioral data include</p> <ul style="list-style-type: none"> • patterns of, or deterrents to, HIV testing 	Refer to Appendix A to locate sources of behavioral data in your service area.

Type of Data	Description	Where to Obtain
	<ul style="list-style-type: none"> • substance use and needle sharing • sexual behavior, including unprotected sex • sexual orientation • health-care-seeking behavior • adherence to prescribed antiretroviral therapies <p>Examples: Supplement to HIV/AIDS Surveillance; HIV Testing Survey; Behavioral Risk Factor Surveillance System; Young Men’s Survey; Survey of HIV Disease and Care Project; Monitoring Trends in Prevalence of STDs, TB, and HIV Risk Behaviors Among Men Who Have Sex with Men Project; Gonococcal Isolate Surveillance Project; CDC HIV Behavioral Surveys; Project One; and the Context of HIV Infection Project</p>	
Clinical data	<p>Clinical data refer to information on the condition(s) of persons with HIV or AIDS. Clinical information is collected so as to understand</p> <ul style="list-style-type: none"> • disease status at the time of diagnosis and later progression (e.g., CD4+ cell count, viral load, opportunistic infections) • type of medical care received • prescription of antiretroviral therapy • type of therapy received <p>Patient surveys collect data on adherence to therapy and health-care-seeking behavior. Depending on the source, clinical data may represent all cases of reported HIV and AIDS or only a fraction. Because clinical data rely on the extent of documentation in a medical record and an ability to locate the record, they may be incomplete.</p> <p>Examples: Adult /Adolescent Spectrum of Disease Project, Survey of HIV Disease and Care Project, AIDS Progression Study, HIV Outpatient Study, and other locally available cohort data</p>	Refer to Appendix A to locate sources of clinical data in your service area.
Demographic data	Demographic data are used to describe social characteristics (e.g., gender, stage of life, and race/ethnicity) of persons in the service area.	Available for state and metropolitan areas from the Bureau of the Census. Also, states maintain census centers. Obtain these data from http://www.census.gov .
Hepatitis B and C surveillance	Data on hepatitis B and C virus (HBV, HCV) infections may represent markers for needle sharing and sexual behaviors, which can be	State health department and CDC staff. The quantity and the quality of surveillance data differ

Type of Data	Description	Where to Obtain
	<p>risk factors for HIV transmission. Data on hepatitis B and C are used to</p> <ul style="list-style-type: none"> • predict the likelihood and rate of spread of viral hepatitis and HIV infections in a community • monitor trends • identify needs for HIV prevention and care services <p>Acute hepatitis B and C (i.e., clinical illness with laboratory confirmation) is reportable in all states; however, because of underreporting and asymptomatic infections, data are likely to be incomplete. More than 40 states have registries for HBV and HCV infection, and most have laboratory reporting laws requiring reporting of positive serologic test results for HBV and HCV infection. Although serologic markers for HBV infection can distinguish between acute and chronic infection, laboratory reports of positive HCV-antibody results cannot differentiate newly acquired infections from chronic or resolved infections, making it difficult to monitor disease trends for HCV. In addition, many registries are relatively new, and their usefulness has not been evaluated.</p>	<p>between states. Refer to CDC's National Notifiable Disease Surveillance System (soon to become the National Electronic Disease Surveillance System) and the CDC Division of Viral Hepatitis.</p> <p>Reference: CDC. <i>Guidelines for Viral Hepatitis Surveillance and Case Management</i>. Atlanta: CDC; 2002. Available at: http://www.cdc.gov/ncidod/diseases/hepatitis/resource/pubs.htm.</p>
Qualitative methods	<p>Qualitative methods are used to obtain data through observations, interviews, discussion groups, focus groups, and analysis of social networks.</p> <p>Example: Rapid Assessment Response and Evaluation project</p>	<p>Health department staff and local community researchers often use qualitative methods to conduct research. Planning group members may also be aware of local studies. Additional information can be obtained from the University of Texas–Southwestern (http://www3.utsouthwestern.edu/preventiontoolbox).</p>
Ryan White CARE Act data reports	<p>The CARE Act data report (CADR) is a form used to collect information annually from grantees and service providers funded under Titles I, II, III, or IV of the Ryan White CARE Act. The CADR collects general information on provider and program characteristics, including the types of organizations providing services (such as ownership status), sources of revenue, expenditures, and paid and volunteer staff. Additionally, the CADR is used to collect aggregate demographic information from which duplicates have been removed (e.g., gender, race, age, HIV exposure category) on total counts of clients served by each provider as well as health insurance coverage and utilization data</p>	<p>Available in all 50 states and all 51 EMAs. Obtain these data from local Ryan White Title I or Title II grantees.</p>

Type of Data	Description	Where to Obtain
	<p>about medical and support services.</p> <p>The CADR is the only source of Ryan White CARE Act data that is available in all states and eligible metropolitan areas (EMAs). It provides demographic information and service utilization data on all Ryan White CARE Act clients. In some areas, Title I or Title II grantees have access to unduplicated data across an entire EMA or state. Because it is a summary report by provider, the CADR cannot be used to generate demographic cross-tabulations.</p>	
Sexually transmitted disease (STD) surveillance	<p>These data are used in reports of notifiable STDs such as syphilis, gonorrhea, chancroid, and chlamydia. Use STD surveillance data to obtain the number of cases and incidence of specific STDs. Demographic and clinical data are available from STD surveillance data. They may serve as a surrogate marker for unsafe sexual practices in a specific risk population. STDs are reportable in all 50 states and US territories. Despite widespread availability, reporting of STDs from private-sector providers may be less complete. Although STDs are the result of unsafe sexual behavior, STDs are not necessarily good predictors of HIV infection.</p>	<p>Available in all 50 states and US territories. Contact the STD program manager in your service area for information.</p>
Socioeconomic data	<p>Socioeconomic data are used to describe characteristics (e.g., income, education, poverty level) of persons in the service area.</p>	<p>Available for state and metropolitan areas from the Bureau of the Census and the Bureau of Labor Statistics. Additionally, states maintain census and labor statistics centers. Obtain these data from http://www.census.gov, http://www.bls.gov, and state census centers.</p>
Special studies and surveys at the local level	<p>Surveys and other data collected from community-based organizations, AIDS service organizations, universities, and special studies. Includes recurring surveys in at-risk populations.</p>	<p>Local researchers and universities</p>

Type of Data	Description	Where to Obtain
Substance abuse data	<p>Substance abuse data are obtained from population-based surveys, medical examiner records, correctional facilities, law enforcement agencies, and drug treatment centers. These sources describe the patterns, prevalence, and consequences of drug use in the general population and specific populations.</p>	<p>National Institutes of Health (http://www.drugabuse.gov/NIDAHome.html); Substance Abuse and Mental Health Services Administration (http://www.samhsa.gov) for information from drug-use surveys and data on treatment and drug abuse; National Institute of Justice (http://www.adam-nij.net) for drug abuse among persons who have been arrested</p>
Tuberculosis surveillance	<p>All 50 states, the District of Columbia, New York City, Puerto Rico, and other US jurisdictions in the Pacific and Caribbean report tuberculosis (TB) cases to CDC on a standard case report form. In 1993, in conjunction with state and local health departments, CDC implemented an expanded surveillance system to collect additional data to</p> <ul style="list-style-type: none"> • better monitor and target groups at risk for TB • estimate and follow the extent of drug-resistant TB • evaluate outcomes of TB cases <p>Although information on HIV status among reported TB cases is available, it may not be complete because of</p> <ul style="list-style-type: none"> • confidentiality concerns that limit the exchange of data between TB and HIV/AIDS programs • local or state laws and regulations that prohibit the HIV/AIDS program and the TB program from sharing information about patients • reluctance of health care providers to report HIV test results to the TB surveillance program staff • a lack of counseling and HIV testing for some TB patients 	<p>Available in all 50 states and US territories. Contact the TB program manager in your service area for information.</p>
Vital records	<p>Vital records contain information, as stipulated by state statutes, on all births and deaths in the 50 US states, the District of Columbia, and US territories. For example, death records include</p> <ul style="list-style-type: none"> • the cause of death according to the rules of the National Center for Health Statistics and the International Classification of Diseases (ICD-9 or ICD-10) • date of death • demographics of the deceased 	<p>All states maintain registries of deaths. Contact the State Vital Records Registrar.</p>

Section 5: Identifying Skills for Preparing Profiles

The goal of the guidelines is to help you produce epidemiologic profiles that are consistent in content, format, usefulness, and quality for prevention and care so as to promote comparability across jurisdictions and the equitable allocation of scarce resources to populations most in need. The guidelines dictate a common set of standards and principles for epidemiologic profiles. The achievement of these goals relies in part on the skills and knowledge of the profile writers and the resources available to them. However, the personnel, expertise, and resources for preparing HIV/AIDS epidemiologic profiles differ by service area.

A Multidisciplinary Approach

A multidisciplinary team approach is recommended, even if one person takes the lead in writing the profile. A team can reduce strain on local resources (i.e., one person who prepares the profile in addition to other responsibilities) and bring multiple skills and experience to the work.

Skills for a multidisciplinary team

A multidisciplinary team should include persons with knowledge and skills in the following areas:

- familiarity with clinical aspects of HIV/AIDS and its treatment
- familiarity with the local HIV/AIDS epidemic
- familiarity with strengths and limitations of available data sources
- skills in data analysis and interpretation
- knowledge of, and experience with, programs and research
- knowledge of the needs and duties of the prevention and care planning groups (setting priorities among populations, interventions, and services)
- understanding of how the epidemiologic profile relates to HRSA and CDC requirements for needs assessment and gap analysis
- knowledge of the data needs of the people carrying out prevention and care programs
- knowledge of policy issues
- knowledge and acceptance of the tenets of applicable confidentiality protocols
- ability to communicate to a diverse audience in user-friendly language

Desired Proficiencies

Whether the profile is being prepared by in-house staff or a consultant(s), certain minimum skills and knowledge are needed to ensure a valid, useful profile. Additional capabilities can enhance the development of the profile.

Minimum knowledge base

- knowledge of HIV/AIDS surveillance systems
- knowledge of basic principles of epidemiology and statistics
- basic knowledge of CDC and HRSA programs in HIV/AIDS prevention and care
- understanding of the confidential nature of HIV/AIDS and other data (e.g., restrictions in reporting small numbers)

- familiarity with the local HIV/AIDS epidemic
- understanding of how data are collected and the strengths and weaknesses of the data
- ability to interpret data from HIV/AIDS and other surveillance systems to make inferences for HIV prevention and care planning

Minimum skills

- statistical skills
 - using descriptive statistics (e.g., mean, median, frequency, percentage, statistical relationships)
 - calculating rates
 - assessing trends over time
- computer skills
 - word processing
 - use of basic statistical and graphics software (e.g., Microsoft Excel, Microsoft PowerPoint, Epi Info, Statistical Package for the Social Sciences, Statistical Analysis Software)
- writing and speaking skills, including the ability to communicate difficult concepts clearly to a variety of end users
- interpersonal skills and ability to work with persons from diverse backgrounds and disciplines

Special Considerations: Working with a Consultant

Some planning groups may wish to employ a consultant to prepare or assist in preparing the profile. For example, a small EMA with limited resources might hire an epidemiologist from a local university to work with the planning groups.

When using a consultant, be sure to observe the following:

- Develop a contract that clearly delineates the scope of work, the timeline, and ownership of the final product (the final document should be in the public domain [i.e., not copyrighted]; the consultant should acknowledge the source when publishing work that includes information from the profile).
- Choose a person with the skills and experience necessary to execute the work defined in the scope of work (some consultants may need help understanding CDC and HRSA requirements, and the added “costs” of ensuring that the consultant is knowledgeable enough to complete the work should be factored into your planning for the time and resources you will need to develop the profile).
- Provide ongoing supervision and oversight of the consultant’s work; build in supervision through such activities as periodic reviews of drafts.
- Ensure that people who have experience with, and expertise in, collecting and working with surveillance data review drafts of the profile.
- Ensure that consultants who need access to, or use, confidential public health surveillance data (e.g., HIV, STD, TB data) be required to adhere to confidentiality and data release restrictions and be subject to penalties for violating these restrictions.

For more guidance on choosing and working with a consultant, see the HRSA HIV/AIDS Bureau, Evaluation Monograph Series Report 1, *Choosing and Using an External Evaluator* (<http://www.hab.hrsa.gov/tools/topics/monographs.htm>).

Section 6: Understanding Basic Analytic Concepts

A common understanding of key terms in data analysis and the methods to be used in developing the HIV/AIDS epidemiologic profile is critical for a planning group. This section presents basic terms and definitions and, when applicable, the methods you are encouraged to adopt when preparing your profile. See the glossary for other relevant terms.

Terms, Definitions, Calculations

case: A condition, such as HIV (e.g., an HIV case) or AIDS (e.g., an AIDS case), according to a standard case definition.

confidence interval (CI): A range of values for a measure that is believed to contain the true value at a specified level of statistical certainty (e.g., 95%).

convenience sampling: A technique that relies upon selecting people who are easily accessible at the time of a survey (e.g., a survey of clients who attend a group meeting or are in a clinic when a researcher happens to be there).

The advantage of convenience sampling is that it is easy to carry out. The weakness is that the findings may not represent the group you are trying to study.

cumulative cases: The total number of cases of a disease reported or diagnosed during a specified time. Cumulative cases can include cases in people who have already died.

Example: Assume that 9,000 AIDS cases had been diagnosed in a state from the beginning of the epidemic through the year 2001. Among the 9,000 persons with AIDS, 4,000 had died. The cumulative number of AIDS cases diagnosed in that state through 2001 would be 9,000.

cumulative incidence rate: The total number of cases during a specified time period, among all people at risk for the disease.

A cumulative incidence rate is calculated by dividing cumulative incidence for a specified time period by the population in which cases occurred during the time period. A multiplier is used to convert the resulting fraction to a number (numerator) over a common denominator, often 100,000.

Number of new cases in specified period

Population at risk in specified period X 100,000

Example: Assume that from 1990 through 2001, 19,000 AIDS cases occurred in a state. During the same time 1,900,000 people lived in the state.

$$\begin{array}{l} \text{Cumulative} \\ \text{incidence rate} \end{array} = \frac{19,000}{1,900,000} \times 100,000 = 1,000 \text{ AIDS cases per } 100,000 \text{ persons}$$

estimate: When accurate data are not available, an estimate may be based on the data that are available and an understanding of how they can be generalized to larger populations. In some instances, national or state data may be statistically adjusted to estimate local conditions. Estimates should be accompanied by statistical estimates of error (a confidence interval), which describe the uncertainty associated with the estimate.

Example: The estimated HIV incidence in State X was 2.1% per year (95% CI, 1.4–2.6).

incidence: The number of new cases in a defined population in a certain time period, often 1 year, which can be used to measure disease frequency. It is important to understand the difference between HIV incidence and reported HIV diagnoses. HIV incidence refers to all persons infected with HIV during a specified period of time (usually 1 year). However, new diagnoses include cases in persons who have been infected for longer periods; they do not include cases in persons who were tested anonymously. Because anonymous test results are not included, HIV surveillance data may not represent incident cases.

Example: During the year 2001, a total of 1,100 AIDS cases were diagnosed in a given state. This is the incidence of AIDS for 2001 in that state.

incidence rate: The number of new cases in a specific area during a specific time period among those at risk in the same area and time period.

Incidence rate provides a measure of the effect of illness relative to the size of the population. Incidence rate is calculated by dividing incidence in the specified period by the population in which cases occurred. A multiplier is used to convert the resulting fraction to a number over a common denominator, often 100,000.

$$\frac{\text{Number of new cases in specified period}}{\text{Population at risk in specified period}} \times 100,000$$

Example: Assume that during the year 2001, a total of 1,100 AIDS cases were diagnosed in a given state. This is the incidence of AIDS for 2001 in that state. The population in the state was 2,200,000 in 2001.

$$\text{The incidence rate} = \frac{1,100}{2,200,000} \times 100,000 = 50 \text{ per } 100,000 \text{ persons in the state}$$

interpretation: The explanation of the meaning of available data. An example is examining a trend, such as the number of HIV cases diagnosed during a 5-year period. Interpreting a trend enables a planning group to assess whether the number of events is increasing or decreasing. However, groups should use caution in interpreting trends that are based on small increases or decreases.

mean: The sum of individual scores in a data set divided by the total number of scores. The mean is what many people refer to as an average.

Example: Assume that people in a given service area in 2001 are the following ages at diagnosis of HIV: 18, 18, 19, 20, 22, 23, 26, 31, 41. The total of the 9 ages = 218 years.

$$\frac{218 \text{ years}}{9} = 24.2 \text{ years}$$

median: The middle value in a data set. Usually, approximately half the values will be higher and half will be lower. The median is useful when a data set contains a few unusually high or unusually low values, which can affect the mean. It is also useful when data are skewed, meaning that most of the values are at one extreme or the other.

Example: Assume the following ages at diagnosis of HIV in the year 2001 data for a given service area: 18, 18, 19, 20, 22, 23, 26, 31, 99. Although the mean age is 30.7, the median age is 22. In this instance, the median age better reflects the central value for age in the population.

no identified risk (NIR): Cases for which epidemiologic follow-up has been conducted, sources of data have been reviewed—which may include an interview with the patient or provider—and no mode of exposure has been identified. Any case that continues to have no reported risk 12 or more months after the report date is considered NIR.

no reported risk (NRR): Cases in which risk information is absent from the initial case report because the information had not been reported by the reporting source, had not been sought, or had not been found by the time the case was reported. Cases may remain NRR until epidemiologic follow-up has been completed and potential risks (exposures) have been identified. If risk has not been identified within 12 months of being reported as NRR, the case may be considered NIR.

percentage: A proportion of the whole, in which the whole is 100.

Example: Assume that 15 of the 60 cases of AIDS in a given year in a state occurred in women.

$$\frac{15}{60} = .25 \times 100 = 25\%$$

prevalence: The total number of cases of a disease in persons not known to have died in a given population at a particular time.

Example: By the end of 2001, if the cumulative number of persons with AIDS in State X is 1,900 and 1,000 have died, then the prevalence of AIDS in State X is 900 (1,900 persons who have ever had a diagnosis of AIDS minus 1,000 who have died).

Prevalence does not indicate how long a person has had a disease and cannot be used to calculate rates of disease. It can provide an estimate of probability that an individual in a population will have a disease at a point in time. For HIV/AIDS surveillance, prevalence refers to persons living with HIV or AIDS regardless of time of infection or diagnosis date. Note the difference between the prevalence of a condition in the population and the prevalence of cases, namely, that a case must be diagnosed according to a definition.

probability sampling: A technique that relies upon random selection to choose individuals from a defined population; all individuals have a known chance of selection. Types of probability samples include simple random sample, systematic random sample, stratified sample, and cluster sample.

proportion: A portion of a complete population or data set, usually expressed as a fraction or percentage of the population or data set.

Example: Assume that 12 of 20 HIV prevention programs in a given service area are school-based programs.

To calculate the proportion as a fraction,

$$\frac{12}{20} = .6 = 6/10 = 3/5$$

To calculate the proportion as a percentage,

$$\frac{12}{20} = .6 \times 100\% = 60\%$$

qualitative data: Information from sources such as narrative behavior studies, focus groups, open-ended interviews, direct observations, ethnographic studies, and documents. Findings from these sources are usually described in terms of common themes and patterns of response rather than numerically or statistically. For the purposes of epidemiologic profiles, qualitative data are useful as supplements to surveillance data to obtain information on risk behaviors and associated factors in specific locales or populations that may not be well represented in routine surveillance data.

quantitative data: Numeric information (e.g., numbers, rates, and percentages).

range: The values of the largest and smallest values in a data set.

Example: Assume the following ages at diagnosis of HIV in the year 2001 in a given service area: 18, 18, 19, 20, 22, 23, 26, 31, and 41. The range is 18–41.

rate: A measure of the frequency of an event or a disease compared to the number of persons at risk for the event or disease. Usually, when rates are being calculated for an epidemiologic profile, the general population, rather than the population potentially exposed to HIV infection by various high-risk behaviors, is used as the denominator. The size of the general population is known from census data, whereas the size of the high-risk population is usually not known.

$$\frac{\text{Number of reported HIV cases occurring during a given period}}{\text{Population at risk during the same period}} \times 100,000$$

For ease of comparison, the multiplier (100,000) is used to convert the resulting fraction to number of cases per 100,000 population. The choice of 100,000, although arbitrary, is standard practice.

Example: Assume that 16 cases of HIV were reported in a service area and that 400,000 persons live in the area.

To calculate the rate,

$$\frac{16}{400,000} \times 100,000 = 4 \text{ per } 100,000$$

This means that 4 of every 100,000 persons at risk have been reported.

sample: A group selected from a total population with the expectation that studying this group will provide relevant information about the total population.

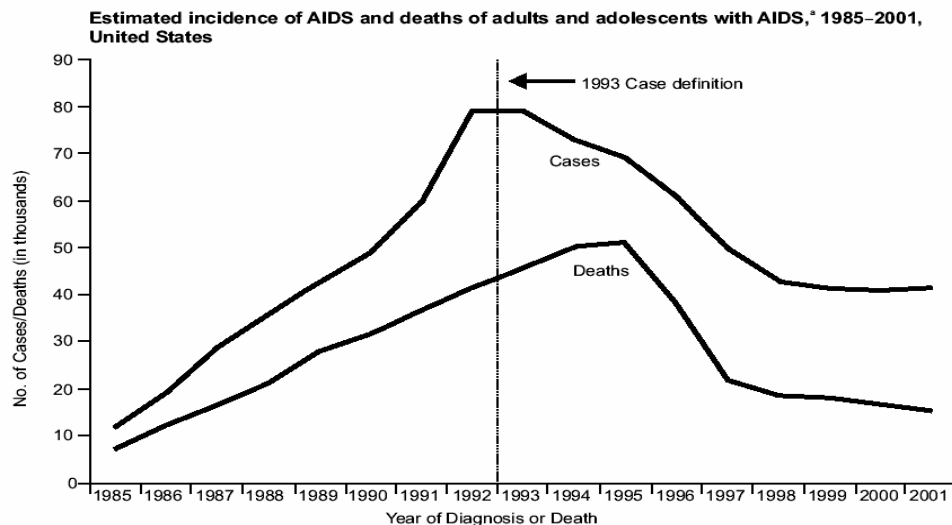
seroprevalence: The number of persons in a defined population who test positive for HIV infection (based on HIV testing of blood specimens). (Seroprevalence is often presented either as a percentage of the total specimens tested or as a rate per 100,000 persons tested.)

stratification: The separation of a sample into subsamples according to predetermined criteria, such as age group, gender, socioeconomic status. Stratification is used to control confounding effects and to detect modifying effects.

trend: A long-term change in frequency, usually an increase or a decrease. A simple linear trend could be described by calculating how much the quantity being measured increased (or decreased) from the beginning value (at the beginning of the period) to the ending value (at the end of the period). The trend could be further described by calculating a time-rate of change in the quantity measured. This is the difference between the beginning and ending values, divided by the number of time units (e.g., years) for which the trend is measured. This calculation yields the amount of increase (or decrease) per time unit. Another key factor is the statistical significance of the trend, which could be a problem if the annual values fluctuate widely from year to year, which would be likely for small numbers.

Trends can be illustrated graphically, by plotting the number of events by time, as shown in Figure 2-2.

Figure 2-2
Example of trend graph



Source. CDC.

^aAdjusted for reporting delays.

Introduction to Analysis and Interpretation

Collecting and presenting HIV/AIDS data are only part of the task. To be useful to planning groups and others, the data must be analyzed and interpreted.

Analysis is the application of logic in order to understand and find meaning in the data. It involves identifying consistent patterns and summarizing the relevant details.

The purposes of analysis in an HIV/AIDS epidemiologic profile are to:

- identify populations that are infected with HIV and describe their key characteristics
- understand the trends and the impact of HIV/AIDS in a service area
- identify groups or populations at risk of acquiring or transmitting HIV and identify their prevention needs
- identify emerging populations and their needs

The following are a few general guidelines for analyzing and interpreting data for the HIV/AIDS epidemiologic profile.

Descriptive analysis

Descriptive analysis is concerned with organizing and summarizing health-related data according to time, place, and person. An example of descriptive analysis might be “The exposure category for 44% of men reported with AIDS in the United States in 1999 was male-to-male sex.”

To carry out an effective descriptive analysis, become familiar with the data before applying analytic techniques. This initial examination should progress to summarizing the data with descriptive statistics, such as frequencies and percentages, in a table to explain the distribution of the HIV/AIDS epidemic in your service area.

As you analyze and interpret your data, keep the following cautions in mind:

- Be aware of the strengths and limitations of the data source. For example,
 - AIDS data do not include those who have been infected most recently.
 - Not all areas report HIV data.
 - EMA service areas may have dissimilar HIV reporting systems (e.g., EMA geographic boundaries cross state lines of 2 states that have different HIV reporting requirements).
- Surveillance data reflect where a person lived when the diagnosis of HIV or AIDS was made, which may or may not be where the person currently lives.
- Confidentiality of public health data is a special concern when dealing with small numbers of cases because of the potential that a person can be identified.
- Interpret surrogate or proxy data with caution (e.g., using STD data as a marker for HIV exposure or infection).

- Concerns about lack of reliability mean that you should be careful about overinterpreting large percent changes (increases or decreases) based on small numbers.

Example: You observe a 200% increase in cases in one group versus a 5% increase in another. However, the 200% increase represents a change from 2 cases in 1999 to 6 cases in 2000; the 5% increase represents a change from 1,000 cases to 1,050 cases. This is an absolute difference of 4 versus an absolute difference of 50. The 200% increase could be due to fluctuations typical of small numbers. Or perhaps 2 of the 6 cases in 2000 should have been reported in 1999. If so, then 4 cases would have been diagnosed in each of the 2 years, and there would have been no increase.

Triangulation

Triangulation, or data synthesis, refers to comparing and contrasting the results of different kinds of research that address the same topic. For example, you may want to see whether the same methods lead to similar findings (e.g., do biologic data and surveys indicate similar patterns in HIV prevalence?). The similarity of results from very different data is referred to as *convergent validity*.

When research findings from different studies or different methods are robust (i.e., not very sensitive to departures from assumptions, for example, that the data are normally distributed), profile writers have an empirical basis for making stronger statements about the validity of their findings and conclusions. If HIV prevalence data, AIDS prevalence data, STD prevalence data, and surveys of risk behavior show consistent evidence of higher HIV risk in a population, then you can be much more confident in saying that this population should be given a high priority for prevention services than you could be if you have only one kind of data. This is why multiple indicators of risk that address different aspects of HIV risk and use different methods are useful. Besides providing another index of validity, convergent findings may be clearer and more convincing to planning group members, service providers, policymakers, and others.

By the same token, different data may suggest contradictory findings. When this occurs, it is important for epidemiologists to account for the reasons that different studies have arrived at different conclusions. This process can be important in terms of identifying problems in data collection or previously undetected differences within populations. Surveys collected under poorly monitored conditions may yield results that are different from those in which the population is well characterized and sampling procedures are rigorously followed. Recent data such as HIV case reporting may reveal emerging populations at risk that are not evident from AIDS case reporting. Survey studies of drug use may suggest that methamphetamine injection may be increasing in a particular population, but no change has yet been seen in HIV prevalence. This may mean that HIV infection has not yet entered the population, which would suggest the need to look specifically at risk practices of this population that have protected them from HIV infection and also look at “mixing patterns” (persons with whom they share drugs and persons with whom they have sex). The use of rapid assessment in such a population

could lead to a better understanding of the epidemiology of a potential new epidemic. Divergent patterns like these also may suggest areas that should be investigated during the prevention needs assessment.

The simplest way to triangulate, or synthesize, data in the profile is to look at the main demographic categories and see how they differ according to data sources. Hence, you may want to look at similarities or differences across data sources by race/ethnicity, gender, geographic area, and age group. Summary statements based on triangulation of the data will be helpful to profile users in understanding how to integrate the large number of tables, figures, and findings that are typically included in an epidemiologic profile.

Where to Get Technical Assistance

If a state or local HIV/AIDS surveillance coordinator is not preparing the profile or is not part of the team preparing the profile, you may want to seek that person's assistance. The HIV/AIDS surveillance coordinator will be able to provide technical assistance in acquiring, analyzing, and interpreting core HIV/AIDS surveillance data. Also consult with the HIV prevention or care programs in the health department about remaining questions or needs for technical assistance.

If your technical needs cannot be addressed at the local level, technical assistance is available both from HRSA and CDC.

For CARE Act grant requirements

For technical assistance needs that relate directly to CARE Act grant requirements, contact HRSA. All technical assistance requests must go through the project officer:

HIV/AIDS Bureau
Division of Service Systems
Health Resources and Services Administration
5600 Fishers Lane, Room 7A-07
Rockville, MD 20857
301-443-9086

For prevention grant requirements

For technical assistance needs that relate to prevention cooperative agreement requirements, contact the Prevention Program Branch at CDC:

Chief, Prevention Program Branch
Division of HIV/AIDS Prevention
National Center for HIV, STD, and TB Prevention
Centers for Disease Control and Prevention
Mailstop E-58
1600 Clifton Road, NE
Atlanta, GA 30333
404-639-5230

For developing epidemiologic profiles for HIV prevention community planning

For technical assistance needs that relate to developing epidemiologic profiles for HIV prevention community planning, contact the HIV Incidence and Case Surveillance Branch at CDC:

Chief, HIV Incidence and Case Surveillance Branch
Division of HIV/AIDS Prevention
National Center for HIV, STD, and TB Prevention
Centers for Disease Control and Prevention
Mailstop E-47
1600 Clifton Road, NE
Atlanta, GA 30333
404-639-2050

Other sources

Other sources of technical assistance include researchers at local universities (such as those at schools of public health, programs in community health and education, and social science departments) and organizational entities, such as the American Psychological Association's Behavioral and Social Scientist Volunteers Program.