

Appendix D Data Verification and Validation

Data verification and validation help to ensure that the data CDC uses to assess performance is of sufficient quality. The following data systems have been referenced in the CDC Performance Plan as sources for data used in assessing program implementation and effectiveness.

Behavioral Risk Factor Surveillance System

In 1984, CDC initiated the Behavioral Risk Factor Surveillance System (BRFSS), a unique, state-based surveillance system designed to collect prevalence data on behavioral risks and conditions that affect health. States conduct monthly telephone surveys using a standardized questionnaire to determine the distribution of behavioral risk factors. Survey responses are forwarded to CDC, where the data are aggregated and published at year's end. The BRFSS provides flexible, timely, and ongoing data collection that allows for state-to-state and state-to-nation comparisons. Participating states use data derived from the BRFSS to identify demographic variations in health-related behaviors, target services, address emerging and critical health issues, propose legislation for health initiatives, and measure progress toward state and national health objectives. The system's broad network of information gathering also enables states to evaluate their disease prevention and health promotion efforts.

The BRFSS survey instrument is a three-part questionnaire developed jointly by CDC and the states:

1. Core component: The *fixed core* is a standard set of questions asked by all states on demographic characteristics and behaviors that affect health (e.g., tobacco use, alcohol consumption). The *rotating core* includes two sets of questions, each asked in alternating years by all states, that address different topics. The *emerging core* consists of up to five questions that typically focus on late-breaking issues. These questions are added to the core for one year and evaluated at year's end to determine their potential value in future surveys.
2. Optional CDC modules: These are sets of questions on specific topics (e.g., smokeless tobacco use, arthritis) that states can opt to include in their questionnaires.
3. State-added questions: These questions are developed or acquired by participating states and added to their questionnaires.

Each year, states and CDC agree on the content of the core components and optional modules. For ease of comparability and use, many of the questions are taken from established national surveys. More than 30 validity and reliability studies attest to the quality and validity of data derived from the BRFSS.

Appendix D Data Verification and Validation

Clinical Laboratory Improvements Act of 1988 (CLIA)

The Clinical Laboratory Improvements Act of 1988 (CLIA) is designed to ensure the sound and scientific development of new laboratory methods. CLIA includes standards that must be met before certification of a laboratory method. These standards include an exacting series of internal and external evaluations. Among the internal checks is the development of a detailed procedures manual for each method. Manuals must be verified and approved by senior laboratory personnel who were not directly involved in the development of the method. CLIA also provides detailed specifications for quality control and calibration of laboratory equipment. Further internal control is provided through regular review from a designated Quality Assurance Officer tasked with ensuring that generally accepted international scientific standards are being followed in the development of the method. External evaluation and control are provided through regular on-site inspections by statutorily approved, independent inspection teams. Inspectors review the internal procedures established by the organization to ensure compliance with CLIA standards. To date, CDC has passed all on-site CLIA inspections.

Group B Streptococcal Disease Surveillance, part of the Active Bacterial Core Surveillance (ABCs)

In 1989, CDC initiated active surveillance for group B streptococcal (GBS) disease as part of the Active Bacterial Core Surveillance (ABCs) system, an active surveillance system for several pathogens that cause invasive disease. Surveillance was conducted in five geographic areas that were awarded contracts after a competitive request for proposals. In 1994, active surveillance for GBS disease was included as a core activity of the newly established Emerging Infections Program (EIP) network, a cooperative agreement program that addresses important public health issues related to infectious diseases. In 1999, the EIP network comprised eight states; all participated in ABCs and conducted active surveillance for invasive GBS disease.

Specific objectives for GBS disease surveillance are to: 1) assess the impact of CDC prevention guidelines published in May 1996, 2) determine the extent to which continuing cases of early-onset GBS disease are preventable through current prevention strategies, 3) identify serotypes responsible for disease to guide vaccine development, 4) evaluate progress in the elimination of serotype b disease, 5) detect possible emergence of disease due to other capsular types, and 6) determine possible preventable reservoirs of the bacteria. Data collection focuses on disease occurrence. State surveillance officers contact personnel in all microbiology laboratories that process bacterial cultures from sterile sites to find cases of GBS. Laboratory audits are also conducted semi-annually to detect possible underreporting. Data are transmitted electronically from the EIPs to CDC's ABCs team on a monthly basis. Annual surveillance reports are made available on the Internet at the ABCs website. Laboratory testing of isolates collected as part of surveillance is performed in reference laboratories. Electronic files containing results of laboratory testing of each state's isolates are fed back to that state on a monthly basis.

Appendix D Data Verification and Validation

Routine laboratory audits to ensure the completeness of data collection represent a tremendous strength of the system. Each month, CDC staff review data and transmit potential errors to state personnel for evaluation. Performance standards for active surveillance have been established in each site to permit aggregation of data collected via somewhat different approaches. Detailed instructions for completion of case report forms ensure consistency across sites. State surveillance officers and CDC's ABCs team hold monthly conference calls to address logistical and technical aspects of the system and meet annually to review and update protocols, present special studies, and discuss innovations. Site visits are currently conducted on an as-needed basis, but annual site visits are planned.

Easy access to the data is provided through a website that includes the basic protocol and one-page yearly surveillance reports for each pathogen. Additional information on GBS is available on a website focused on that infection, with many materials targeted to pregnant women or healthcare providers and public health workers concerned with pregnant women.

The principal limitation of GBS disease surveillance through the ABCs is that it is not conducted throughout the United States. Substantial geographic variation in the incidence of invasive GBS disease has been noted, and it is unclear whether states outside ABCs areas have experienced changes in the incidence of GBS disease that are comparable to those noted in the surveillance areas. One way of addressing this limitation is to increase the availability of ABCs methods and tools. Through the website and frequent publications, CDC is attempting to provide other state health departments with information that can help them assess whether the efforts involved in conducting invasive GBS disease surveillance, particularly for early-onset disease in infants <7 days, are feasible in their locales.

Integrated Resources Information System

CDC's Integrated Resources Information System (IRIS) is a collection of applications to assist management in budget, staffing, and project planning, tracking, and reporting. The IRIS budget application provides detailed budget information by CDC component. It allows managers to view budget reports grouped by a variety of options. IRIS staffing is a view-only application designed to allow users to quickly access personnel data reports and project employee salaries for a specified time period. The projects application allows managers to plan, track, and manage various types of projects. This application provides access to project data, resources, and administrative functions. All information for a project must be maintained in the IRIS projects component to ensure consistency and reliability of data. The IRIS reports application is the data retrieval and reporting component.

Appendix D Data Verification and Validation

National Health and Nutrition Examination Survey

The National Health and Nutrition Examination Survey (NHANES) is a program of studies to assess the health and nutritional status of adults and children in the United States. Started in the early 1960s, NHANES is the only national source of objectively measured health data capable of providing accurate estimates of both diagnosed and undiagnosed medical conditions in the population. Findings from the survey are essential for determining rates of major diseases and health conditions and for developing public health policies and prevention interventions. The survey screens 15,000 households per year and selects 3,500. From this sample, 5,000 persons are interviewed and examined annually. Samples are recruited from 15 counties or clusters of counties each year. Samples comprise sufficient numbers to provide reliable estimates by gender and age group for non-Hispanic whites, Mexican Americans, and African Americans.

Data are collected via health interview, physical examination, and clinical and laboratory tests. Interviews are conducted in respondents' homes. Physical examinations are performed in specially designed mobile examination centers that travel to survey locations throughout the country. These centers allow for the collection of data on chronic conditions, nutritional status, medical risk factors, dental health, vision, illicit drug use, blood lead levels, food safety, and other factors that are not possible to assess by use of interviews alone. The medical team consists of a physician, dentist, medical and health technicians, and dietary and health interviewers; trained bilingual staff conduct the household interviews.

An advanced computer system using high-end servers, desktop PCs, and wide-area networking is used to collect and process all NHANES data, nearly eliminating the need for paper forms and manual coding operations. Household interviewers use notebook computers with electronic pens for data collection in the field. Data collected in the mobile examination centers are automatically transmitted via a frame relay network into central databases. Survey information is available to CDC within 24 hours of collection.

Information from NHANES is disseminated through an extensive series of publications and articles in scientific and technical journals. Survey data are also available on CD-ROM and computer diskettes. In previous years, data were available for analysis approximately 31 months after collection. A goal is to improve the timeliness of data dissemination. The computerized system has already substantially improved access to the data from the field.

A comprehensive quality assurance program is instituted before data collection begins, with appropriate training that requires significant practice time for the health examiners and interviewers. Training focuses on hands-on experience rather than didactic methods. During data collection, health examiners and survey staff meet regularly to discuss operations, updates, and problems. Staff are retrained as needed.

Appendix D

Data Verification and Validation

NHANES relies on both passive and active monitoring systems for operational and content-related quality control. Passive quality control uses automated computer procedures for detecting data anomalies. After careful analysis, appropriate activities can be undertaken to resolve any data collection issues. Active quality control relies on examiner feedback to identify and evaluate problems and select remedies. NHANES primarily relies on physical measurements from well-established biomedical procedures. In most instances, these measurements represent the gold standard data against which self-reported data might be validated for other subjective data collection modalities. New technologies under consideration are evaluated to determine if they provide valid estimates of the condition, risk factor, or measurement for which they are being used. The evaluation might include a scientific literature review, expert workshop, or validity study.

National Health Interview Survey

The National Health Interview Survey (NHIS) is the principal source of information on the health of the civilian, non-institutionalized population of the United States. The purpose of the NHIS is to monitor the health of the U.S. population through the collection and analysis of data on a broad range of health topics. A strength of the survey is the ability to display these health characteristics by many demographic and socioeconomic factors. NHIS data are used widely throughout DHHS to monitor trends in illness and disability and to track progress toward achieving national health objectives. The data are also used by the public health research community for epidemiologic and policy analysis.

The NHIS is a cross-sectional household interview survey. Sampling and interviewing are continuous throughout each year. Households chosen for interviews are a probability sample representative of the target population. NHIS data are collected annually from approximately 43,000 households including about 106,000 persons. Survey participation is voluntary, and the confidentiality of responses is ensured. The annual response rate is >90% of eligible households in the sample.

The NHIS has three modules:

- The basic module remains largely unchanged from year to year and allows for trend analysis. Data from more than one year can also be pooled to increase the sample size for analytic purposes. The basic module contains a family core, a sample adult core, and a child core through which data are collected on the family unit and from one randomly selected adult and child.
- Periodic modules collect more detailed information on some of the topics included in the basic module.
- Topical modules respond to new data needs as they arise.

Appendix D Data Verification and Validation

Data are collected through a personal household interview conducted by staff employed and trained by the U.S. Bureau of the Census according to procedures delineated by CDC. Data are reviewed and analyzed extensively to ensure their validity and reliability. The survey sample is designed to yield estimates that are representative and that have acceptably small variations.

Before the actual survey, cognitive testing is performed by CDC's Questionnaire Design Research laboratory, and pretests are conducted in the field. Once collected, data are carefully edited, checked, and compared to data from earlier surveys and/or independent sources. Staff members calculate descriptive statistics and perform in-depth analyses, which result in feedback on the analytic usefulness of the data.

In the past, it has taken approximately 26 months for the survey data to be released for a given year. Improving the timeliness of NHIS data is a GPRA performance measure.

National Hospital Discharge Survey

The National Hospital Discharge Survey (NHDS), conducted annually since 1965, is a national probability survey designed to meet the need for information on characteristics of inpatients discharged from non-federal, short-stay hospitals in the United States. The NHDS collects data from a sample of approximately 300,000 inpatient records acquired from a national sample of about 500 hospitals. The NHDS provides national and regional estimates of U.S. inpatient hospital utilization by the demographic characteristics of patients discharged, conditions diagnosed, and surgical and non-surgical procedures performed. Approximately 95% of eligible sample hospitals respond to the survey.

The NHDS uses two data collection methods: 1) a manual system in which hospital staff or staff of the U.S. Bureau of the Census abstract data from medical records, and 2) an automated system in which CDC purchases machine-readable medical record data from commercial organizations, state data systems, hospitals, or hospital associations. Approximately 40% of hospitals provide data through the automated system. Data are generally available about 17 months after collection. Timeliness is being addressed as part of the GPRA effort.

An ongoing quality control program helps to ensure the accuracy of NHDS data. NHDS data have been found to be a good reflection of information found in medical records. What is not known is the degree to which medical record information reflects actual performance.

Appendix D Data Verification and Validation

National Immunization Survey

The Childhood Immunization Initiative (CII) is one of many federal, state, and local programs mounted to raise vaccination levels in young children. The CII established a 1996 goal of increasing vaccination levels for 2-year-old children to at least 90% for measles-mumps-rubella, diphtheria and tetanus toxoids and pertussis vaccine, oral poliovirus vaccine, and *Haemophilus influenzae* type b vaccine. In addition, the CII established a goal for 1996 to increase vaccination levels for 2-year-old children to at least 70% for three or more doses of hepatitis B vaccine.

The National Immunization Survey (NIS) is used to assess progress towards these goals. NIS data provide current, population-based, state and local estimates of vaccination coverage produced by a standard methodology. Quarterly data are collected via household interviews in 50 states, the District of Columbia, and 27 urban areas. Interviews are conducted by telephone with randomly selected households. Each quarter, CDC calculates estimates of vaccination coverage levels and makes valid comparisons of state efforts to deliver vaccination services. CDC uses NIS data to evaluate progress towards national vaccination goals and to identify states with the highest and lowest immunization rates.

To ensure the accuracy and precision of coverage estimates, immunization data for surveyed children are also collected through a mail survey of their pediatricians, family physicians, and other healthcare providers. The parents and guardians of NIS-eligible children are asked during the telephone interview for consent to contact childrens' medical providers. Types of immunizations, dates of administration, and additional data about facility characteristics are requested from immunization providers identified during the telephone survey of households. NIS estimates of vaccination coverage therefore reflect a comparison of information provided by both immunization providers and households.

National Vital Statistics System

Vital statistics are often the most complete and continuous information available to public health officials at the national, state, and local levels. The National Vital Statistics System is responsible for the nation's official vital statistics. The registration of vital events – births, deaths, marriages, divorces, fetal deaths – is a state function, and vital statistics are provided through state-based registration systems. Since 1902, the federal government has obtained use of the records for statistical purposes through cooperative arrangements with the responsible agencies in each state. Standard forms for the collection of data and model procedures for the uniform registration of events are developed and recommended for state use through cooperative activities of the states and CDC. CDC also provides training and instructional materials to the states as part of ongoing technical assistance.

Appendix D

Data Verification and Validation

The purpose of collecting the data is to monitor trends over time through vital life events. Vital records and reports originate with private citizens, such as the family affected by the events, physicians, or funeral directors. By law, birth registration is the direct responsibility of the hospital of birth or the attendant at the birth. In the absence of an attendant, the parents of the child are responsible for registering the birth. Although procedures vary from hospital to hospital, personal information is usually obtained from the mother; medical information may be obtained from the chart or from a worksheet completed by the birth attendant. Reporting requirements vary from state to state; in general, the completed certificate must be filed with the state or local registrar within 10 days of birth. Published data represent all counties and places of 10,000 or more population. Electronic files include data for states, counties, large cities (population of 100,000 or more), and metropolitan statistical areas.

By law, death registration is the direct responsibility of the funeral director or person acting as such. The funeral director obtains the data required, other than the cause of death, from the decedent's family or other informant. The attending physician provides a best medical opinion about the cause and manner of death; later this information is coded by the state or CDC according to uniform codes. Demographic information is also recorded. If no physician was in attendance or if the death was due to other than natural causes, the medical examiner or coroner investigates the death and provides the cause and manner. Reporting requirements for death vary, but in general the completed certificate must be filed within 3 to 5 days of the death. Published data include all counties and places of 10,000 or more population. Electronic files include data for states, counties, large cities (population of 100,000 or more), and metropolitan statistical areas.

Fetal deaths are also reported through the National Vital Statistics System. All fetal deaths of 20 weeks or more gestation that occur in the United States are recorded. A linked birth/infant death file allows for the analysis of demographic and health characteristics from certificates of live births in combination with causes of death and other data from death certificates of infants who died before their first year of life. The linked file set includes information on all the infants who died in the United States each year, as well as information on all live births. An additional file includes information on death records not linked to birth certificates. The match rate is about 97%-98%. Data are organized by calendar year.

Provisional and final estimates of the number of marriages and divorces are obtained from each state able to provide these figures. Since data are not available from all states, national divorce rates are not produced. Detailed characteristics of marriages and divorces have not been available since 1996.

Vital statistics data are collected using uniform procedures and are accurate and consistent. The data are reported as soon as they are analyzed by CDC staff. Monthly provisional numbers and rates are published in the *National Vital Statistics Reports*. These figures are based on approximate counts of the number of events that occurred in a given state; an estimation procedure is used to convert these occurrence estimates into state-specific estimates of the number and rate of resident events. Preliminary data collected through the National Vital Statistics System are made available to the public approximately 10 months after the end of the collection year. Data are presented for a 12-month period and are published semi-annually in the *National Vital Statistics Reports*. Final data are released about 18 months after collection via *National Vital Statistics Reports*, public use data tapes, CD-ROM, Series Reports, the Internet, and journal articles. Use of electronic products have greatly increased the accessibility of the data and reduced the costs to researchers and other users.

Appendix D

Data Verification and Validation

The data collected through the National Vital Statistics System represent all registered vital events in the United States and adequately represent the true rates of events. To more accurately record birth and death information, new birth and death certificates are being designed through a collaborative effort with states, researchers, and other interested parties. The revised certificates reflect changing data needs and emerging public health applications; they will be implemented in 2003.

Sentinel Surveillance for Chronic Hepatitis C

Although a large number of persons in the United States are chronically infected with HCV and many will develop chronic liver disease, the burden of disease has not been well characterized. There is no ongoing surveillance, and few population-based studies have been conducted from which to determine the incidence and prevalence of chronic liver disease and the relative proportion of cases attributable to viral hepatitis and other etiologies. To begin to collect this information, CDC established a pilot surveillance system for chronic liver disease in 1998. The data-collection system has three components:

- A standard interview questionnaire, developed by CDC, is used by all sites to ensure comparability of data and facilitate aggregation of data as appropriate. The instrument includes questions from other established surveillance systems and from previous studies of chronic liver disease. Questions focus on demographic characteristics, clinical information, quality of life issues, and exposures and risk factors.
- A standard form is used to abstract clinical and laboratory information from the patient's clinical chart. This information, collected consistently across sites, includes data needed to determine disease etiology, treatment history, medication use, and other relevant clinical information.
- A serum sample is collected and sent to CDC to identify serologic markers for viral hepatitis.

An important characteristic of the pilot is its comprehensiveness. For the first time, all patients with chronic liver disease in several geographic areas are being identified using a common methodology, with consistent information collected in all sites. The goal is to expand the use of the methodology and data collection instruments to other sites throughout the United States to develop a comprehensive picture of the occurrence and characteristics of chronic liver disease and to monitor trends.

Although quality assurance and quality control instruments are still under development, several validation studies have been conducted. To assess the completeness of reporting, CDC conducted a survey of primary care practitioners and a review of all first-time liver biopsies. These studies indicated that overall surveillance was comprehensive and was successful in identifying the vast majority of patients in the target population. A review of a randomly selected subset of charts failed to reveal any significant errors in chart abstraction. To assess the overall validity of the study, early preliminary results have been compared to the few existing relevant data. This evaluation, demonstrating that the incidence of newly diagnosed chronic liver disease has increased in recent years, is already contributing to CDC's efforts to more accurately estimate the burden of illness from chronic liver disease.

Appendix D Data Verification and Validation

U.S. Sentinel Physician Surveillance for Influenza

Established in 1982, the U.S. Sentinel Physician Surveillance for Influenza is one of four primary sources of influenza surveillance data. The sentinel physician surveillance system is an active system of surveillance conducted from October through May. Each week during that period, several hundred volunteer physicians around the country report the total number of patients seen and the number of those patients with influenza-like illness by age group.

During the 1997-98 influenza season, 27 states and the District of Columbia elected to participate in a pilot program to upgrade the sentinel physician surveillance system. The pilot merged CDC's national sentinel surveillance system and state-based systems into one integrated system based on common methodologies and standards. During the 1998-99 influenza season, the enhanced sentinel physician surveillance system was expanded to include 40 states and the District of Columbia, and an Internet reporting system was developed. States are responsible for establishing, recruiting, and maintaining state-based sentinel physician groups and for ensuring that data are collected and transmitted regularly to a central data repository at CDC, which is updated daily. CDC is responsible for coordinating the system nationally, maintaining the reporting systems, processing and analyzing the data, and maintaining the Internet site. Efforts to improve the system are continuous.

Sentinel physicians can report data via any of three methods: 1) Internet reporting, 2) touchtone phone reporting, or 3) facsimile transmission with manual entry of data. A program developed by CDC integrates the three sources of data and uploads the data to the Internet site. Data are available daily to each state coordinator. A summary of influenza activity is available to the general public each week.

CDC has undertaken a continuous process to simplify use of the system, clarify case definitions, and offer multiple options for input and access. With daily updates and weekly summaries, the information is extremely timely and pertinent for decision making. CDC epidemiologists analyze the data for outlying information and perform routine checks for coherence. State coordinators routinely check the timeliness of reporting and troubleshoot problems at the local level. Guidelines are provided to sentinel physicians for optimal timing of specimen collection for virologic testing on certain patients. There is no way to ascertain that the data on influenza-like illness is free of error, but, as the number of participating sentinel physicians increases, the potential consequences of errors decrease. Given that sentinel surveillance provides an index of current influenza activity, consistent reporting by a stable group of physicians is imperative for data reliability. Increasing sentinel physician sites and sentinel physician participation in each state would greatly increase the validity of the data.

Appendix D Data Verification and Validation

Youth Risk Behavior Surveillance System

CDC established the Youth Risk Behavior Surveillance System (YRBSS) in 1990. One of the components is a national school-based survey that was first conducted in 1990 and has been repeated biennially since 1991. The national Youth Risk Behavior Survey (YRBS) measures six categories of priority health risk behaviors that contribute to the leading causes of mortality and morbidity among youth and adults in the United States: 1) behaviors that may lead to violence and unintentional and intentional injuries; 2) tobacco use; 3) alcohol and other drug use; 4) sexual behaviors that contribute to HIV infection, other sexually transmitted diseases and unintended pregnancy; 5) unhealthy dietary behaviors; and 6) inadequate physical activity.

The YRBS is administered in the spring to nationally representative samples of students in grades 9-12 attending both public and private schools. Professional data collectors, trained specifically for the YRBS, are used as field staff to ensure standard administration procedures. The YRBSS uses a three-stage cluster sample to select schools and classes of students within schools. African-American and Hispanic students are oversampled to provide accurate estimates for these subgroups in each survey cycle. By combining data from multiple survey cycles it is also possible to obtain accurate estimates for Asian and Native American youth. The sample size totals approximately 14,000 students per survey. School response rates average 76%; student response rates average 88%.

The YRBS questionnaire is designed for self-administration by use of a computer-scannable booklet. The questionnaire has been modified as needed to address emerging public health problems. A reliability study of the questionnaire conducted in 1993 demonstrated that students reported health risk behaviors reliably over time. Psychometric work has demonstrated that the questionnaire yields accurate and high-quality data. Standardized data editing and cleaning procedures improve data accuracy and consistency. Data are released within 12 months of data collection and are made available to the public via the Internet.

Appendix D Data Verification and Validation

CDC Program-Specific Data Verification and Validation

Birth Defects, Developmental Disabilities Prevention, and Disabilities and Health

For the goal to prevent birth defects and developmental disabilities, the performance measures use data from CDC's Behavioral Risk Factor Surveillance System, the National Birth Defects Prevention Network, the number of maternal interviews entered into the National Birth Defects Prevention Study, the Alliance for Research in Child Health Epidemiology, and the count of specific types of studies funded by CDC.

For the goal to improve the health and quality of life of Americans with disabilities, the performance measures are simple counts of programs, publications, and data from a database maintained by the Directors of Speech and Hearing Programs for State Health and Welfare Agencies.

Chronic Disease Prevention and Health Promotion

Early Detection of Breast and Cervical Cancer: CDC uses the Minimum Data Elements (MDEs) to report on all GPRA measures. States, territories, and tribal organizations (NBCCEDP grantees) submit MDEs electronically twice a year (January 15 and July 15) to a data management contractor, who analyzes the data and submits a data file to CDC. These files are made available in April and October. CDC will use the January 15 submission to report performance for the new GPRA measures. Data provided in the performance report include only screening exams through March 31 of the previous year to allow adequate time to gather the data and present a complete program report. NBCCEDP grantees are provided 9½ months after the initial screening date (March 31) to gather diagnostic and treatment information and prepare the data submission by January 15. The data management contractor analyzes the data by March and sends the report to CDC. All data collected and submitted by NBCCEDP grantees have indicators to assess completeness. Data are also assessed against established clinical standards.

Tobacco: CDC monitors cigarette use among youth and reports performance on a biennial basis using the Youth Risk Behavior Survey (YRBS), which is a component of the YRBSS (see Appendix A.2). Three additional surveys, the National Household Survey on Drug Abuse (NHSDA) the Monitoring The Future (MTF) Survey, and the National Youth Tobacco Survey (NYTS), provide complementary data for examining trends and understanding youth-related tobacco issues. The NHSDA is conducted annually by SAMHSA; the MTF is conducted annually by the University of Michigan's Institute for Social Research; and the NYTS is currently conducted by the American Legacy Foundation, but will transfer to CDC in 2004.

Community-Based Prevention Research: Data are available from grantee progress reports and will be verified through site visits and publications. CDC program consultants validate information received through site visits and telephone consultations. No data lags are expected.

Heart Disease & Stroke: CDC will evaluate stroke registry capacity via annual state reports, deaths from heart disease and stroke via death certificate data from states, and uncontrolled high blood pressure data from HRSA..

Appendix D Data Verification and Validation

Diabetes: CDC verifies performance through quarterly state reports and periodic site visits. For efforts in American Indian/Alaskan Native populations, data are verified via program reports and documentation of support. The BRFSS collects data on receipt of annual eye and foot exams in persons with diabetes.

Arthritis: CDC collects and evaluates data on state-based arthritis programs via annual state program reports and site visits.

National Cancer Registries: Participating states are expected to collect information on at least 95% of cancer cases diagnosed or treated in their state each year. NPCR funded states are required to incorporate NAACCR standards for data quality and format. States report de-identified cancer case data annually to a CDC contractor. In addition, CDC receives regular reports from each state which summarize progress of completeness, timeliness, and quality of registry data. NPCR staff also prepare annual internal evaluations of program progress.

Variations in states' capacities (planning or enhancement status) and initial funding year result in differences across reference years used for calculating registry data completeness. NAACCR has established a process by which states can apply for certification to ensure that member registries are collecting useful and high-quality data. Member registries are evaluated yearly and provided confidential feedback. Data for FY 2001 will be available in June 2002 for reporting.

HIV Prevention among School-aged Youth: Data are collected on a biennial basis (during odd-numbered years) through CDC's YRBSS, a system designed to focus attention on priority behaviors among youth that are associated with the most important health problems (see Appendix B). The YRBSS was developed in partnership with federal agencies, state departments of education, scientific experts, and survey research specialists. The YRBSS includes separate national, state, and local school-based surveys of high school students. A recent study provides evidence that this adolescent survey has good reliability in measuring health behavior. Baseline data from the 1995 YRBSS are used because: 1) they were the most recent data available when the original measures were created, and 2) they will allow a more accurate illustration of trends in sexual behaviors over time.

Nutrition/Physical Activity and Obesity: CDC plans to collect and evaluate state data on nutrition and physical activity programs via annual state program reports, site visit reports, and a program evaluation database.

Environmental Health

Environmental Health Laboratory/Biomonitoring: All analytical methods developed must be certified under the Clinical Laboratory Improvements Act of 1988 (CLIA).

Appendix D Data Verification and Validation

Data systems at CDC's Environmental Health Laboratory monitor laboratory performance under CLIA. CDC also conducts quality assurance activities internally to confirm results and ensure their validity. CLIA-approved methods are used to analyze levels of environmental chemicals published in the *National Report on Human Exposure to Environmental Chemicals* that are measured in specimens obtained from the National Health and Nutrition Examination Survey (NHANES). The use of CLIA-approved methods is verified by senior staff as well as by internal quality assurance officers. The sample size and control mechanisms for the *Report* have been established as part of NHANES.

Asthma: Data verification is based on required reporting by grantees. CDC project officers will verify that states are fulfilling the requirements of cooperative agreements through routine monitoring of the grants process. CDC epidemiologists will review all statistical and surveillance data to ensure appropriate application of statistical and epidemiologic methods.

Health Statistics

CDC will verify performance via contractor reports, pretest reports, meeting proceedings, publications, and website records.

HIV, STD, and TB Prevention

HIV/AIDS Data Collection Systems: CDC uses multiple data collection systems to monitor HIV trends and prevention programs. The HIV/AIDS Reporting System (HARS) collects case reports of HIV-infected persons in state and local health departments. AIDS case data are available from all states and territories using uniform name-based collection methods (—no names or personal identifiers are sent to CDC; these are maintained only at the local level). Although completeness of reporting of diagnosed AIDS cases varies by area and patient population, studies indicate that reporting in most areas is more than 85% complete. Reporting of AIDS deaths is estimated to be more than 90% complete. In contrast, HIV data collection systems vary between areas (e.g., name-based code, coded identifier, name-to-code, etc. data collection systems). CDC is conducting validation and evaluation studies of these systems to determine the quality of data generated by them. Currently, trends in HIV diagnoses for adults and adolescents are available only from 25 states which have implemented name-based HIV case reporting (using methods similar to those for AIDS case reporting) since at least 1994.

The period of time between a diagnosis of HIV or AIDS and the arrival of a case report at CDC is called the "reporting delay" (40% of AIDS cases are reported to CDC within 3 months of diagnosis, 80% within 1 year). In order to provide the best estimates of trends in incidence, HIV and AIDS surveillance data are analyzed by the date of diagnosis and are mathematically adjusted in more recent periods to adjust for reporting delays and incomplete information on some cases. CDC requires a minimum of 18 months after the end of a calendar year to provide accurate estimates of trends for through that year. For example, calendar year 2000 data will be available in the summer of 2002.

Appendix D Data Verification and Validation

In addition to the HARS data, CDC has supplemental surveillance systems to collect in depth information on HIV/AIDS cases and prevention programs. The Supplement to HIV/AIDS Surveillance (SHAS) project collects interview information from recently reported HIV/AIDS cases ≥ 18 years of age in 16 state/local health department jurisdictions on their sex and drug using behaviors, access to and adherence to care, and utilization of prevention interventions. The Adult and Adolescent Spectrum of HIV Disease (ASD) study collects longitudinal medical record review data on antiretroviral therapy, clinical care, and outcomes from HIV-infected persons receiving care in selected medical facilities in 9 areas; most of these facilities are publicly-funded. The HIV Counseling and Testing System (CTS) collects the number of tests performed, demographic and characteristics, test results, and utilization of post test counseling services in publicly-funded sites in all states.

Surveillance reports and in depth analyses of data from these systems are available upon request from CDC.

Sexually Transmitted Diseases: TD incidence and prevalence data (hardcopy and electronic) undergo ongoing verification and validation procedures including quarterly reports back to project areas comparing reporting across all data sources, trend information, percent unknowns for clinical fields, edit checks and updates, as well as constant communication via fax, phone, and email with project staff. PID hospitalization data is collected through the National Hospital Discharge Survey conducted by the National Center for Health Statistics, and PID initial visits to physicians is collected through the National Diagnostic and Therapeutic Index by IMS America, Ltd. Additional feedback is provided to project areas via annual publications and reports.

Prevention of STD-Related Infertility: Data on the prevalence of chlamydial infection in defined populations have been useful in monitoring disease burden and guiding screening programs. In particular, CDC monitors trends in prevalence among women enrolled in the U.S. Department of Labor National Job Training Program and among women screened for chlamydia attending family planning clinics. These programs provide crucial information on the prevalence of chlamydia in high-risk populations, i.e., young sexually active women. Data from these programs indicate that: 1) chlamydia is geographically widespread (in nearly all states, chlamydia positivity exceeded the Healthy People 2010 objective of 3%), and 2) younger women (<24 years of age) consistently have higher chlamydia positivity than older women. Chlamydia screening is not as widespread for men. Chlamydia prevalence was 4.7% among men aged 17-37 years who were screened at entry in the U.S. Army in 1999-2000. Although these prevalence data are not entirely comparable because of differences in the performance characteristics of screening tests and variations in screening criteria, they provide important information on the continuing high burden of disease. The data also allows monitoring of chlamydia in multiple venues and populations which is critical to understanding the true burden of disease.

In efforts to reduce the prevalence of chlamydia among high-risk women under age 25, CDC does not have activities targeted specifically to Job Training Program participants. However, CDC includes data provided by the U. S. Department of Labor because the data are an important component of assessing burden of disease. National Job Training Program participants, who are required to be screened for chlamydia at program entry, represent an important high-risk population CDC is trying to reach, young sexually active women. Continued expansion of chlamydia screening should lead to a continued reduction of the burden of disease among women, including National Job Training Program participants. For economically

Appendix D Data Verification and Validation

disadvantaged women aged 16 to 24 years who entered the National Job Training Program from 27 states, and Puerto Rico, in 2001, the overall prevalence was 10.6 %. Given that there has been little or no change in the prevalence of chlamydia among the National Job Training Program participants, and given that CDC does not have activities specifically targeting the National Job Training Program, the target has been adjusted to 10% for 2002 and 9% for 2003 and 2004.

In 2001, CDC achieved the goal of reducing chlamydia prevalence among women attending family planning clinics. Unlike the measure that utilizes data from the U.S. Department of Labor's National Job Training Program, this measure reflects the performance of long-standing, widespread CDC-supported screening programs. The median state-specific positivity was 5.6% for women aged 15 to 24 years screened at selected family planning clinics in all states and outlying areas. In selected prenatal clinics in 22 states and Puerto Rico, the chlamydia prevalence was 7.4%. After adjusting trends in chlamydia positivity to account for changes in laboratory methods and associated increases in test sensitivity, chlamydia test positivity among women decreased in five of 10 DHHS regions from 2000 to 2001, increased in four regions, and remained the same in one region. Although chlamydia positivity has declined in the past year in some regions, continued expansion of screening programs to populations with higher prevalence of disease may have contributed to the increases in positivity seen in other regions.

As CDC continues to expand its efforts, data from the family planning clinics is crucial not only in measuring performance but also in guiding future efforts. Effective interventions have been demonstrated, but they are not reaching all those in need. Achieving future declines in chlamydia prevalence hinges upon efforts to: 1) expand chlamydia screening and treatment services so they are easily available to both men and women; 2) increase awareness about chlamydia testing and treatment services at private clinics and doctors' offices; and 3) expand health promotion activities.

Gonorrhea: The U.S. experienced a 73.9% decline in the reported rate of gonorrhea in the U.S. from 1975 to 1997. The rate increased in 1998, but the rates of reported gonococcal infections have since been steady (128.5 in 2001, 129.0 in 2000, 132.3 in 1999, and 131.9 in 1998). The 2001 rate exceeds the Healthy People 2010 objective of 19 cases per 100,000 persons.

Although reported rates of gonorrhea were once substantially higher among men than women, that gap has narrowed. This is most likely due to increased screening in women. Because women are more likely to be asymptomatic than men, cases in women are less likely to be identified and reported. The overall gonorrhea rate in U.S. females in 2001 was similar to the rate in 2000 (128.2 and 126.7, respectively). The gonorrhea rate in men was similar with 130.9 and 128.4 cases per 100,000 males in 2000 and 2001, respectively. Among women aged 15-44, the 2001 rate was 286 per 100,000, exceeding the target rate of 250. In 2001, 15- to 19-year-olds had the highest rate (703.2 cases per 100,000 females) of gonorrhea among women. Among men, rates (563.8 cases/100,000 males) were highest among 20- to 24-year-olds. Profound racial disparities persist for gonorrhea, with 2001 reported rates among non-Hispanic blacks about 27 times higher than among whites and Hispanic rates almost 3 times higher than rates among whites. This disparity most likely reflects differences in access to prevention and treatment services.

Appendix D Data Verification and Validation

Although increased screening, use of more sensitive diagnostic tests, and improved reporting may account for a portion of increase in the recent past, true increases in disease in some populations and geographic areas also appear to have occurred. The southern states continue to have the highest gonorrhea rates of any region. Reasons may include poverty levels and access to quality healthcare and preventive services. Future declines in gonorrhea prevalence will require efforts to 1) increase public and provider awareness of the problem, 2) increase screening and treatment in high-risk populations, and 3) expand health promotion and prevention.

Pelvic Inflammatory Disease (PID): The decrease in the incidence of PID is possible evidence of intensified nationwide screening and treatment efforts for chlamydia, a principal cause of PID. The incidence of hospitalization for PID among women aged 15-44 decreased from 127 per 100,000 women in 1999 to 120 per 100,000 women in 2000, achieving the 2000 target of 125 per 100,000 women. These decreases in hospitalizations may also be attributable to an increasing trend of outpatient management for PID and increased use of oral treatments.

The reported number of initial visits to physicians' offices for PID through the National Disease and Therapeutic Index (NDTI) has generally declined from 1993 through 2001 but is still higher than the 2001 target of <225,000 visits.

Accurate estimates of PID and tubal factor infertility from gonococcal and chlamydia infections are difficult to obtain. Definitive diagnosis of these conditions often requires complex surgical or other diagnostic tests. Most cases of PID are treated on the basis of interpretations of clinical findings, which vary among practitioners. In addition, the settings in which care is provided can vary considerably over time. For example, women with PID who would have been hospitalized in the 1980s may be treated in outpatient facilities today. Future declines in the incidence PID will hinge in part upon expansion of screening and treatment programs for chlamydia and gonorrhea as well as expansion of health promotion efforts that increase both public and provider awareness.

Syphilis Elimination: Syphilis is extremely concentrated geographically. Approximately 80% of U.S. counties have already eliminated syphilis, and 94% have a syphilis rate of ≤ 4 per 100,000. Over 50% of syphilis cases in 2001 were reported from 21 counties. Syphilis remains an important problem in the South and in some urban areas in other regions of the country.

Although provisional data from 2001 indicates continued progress, syphilis elimination efforts are challenged by increases among MSM in areas throughout the country. For example, a gradual increase in syphilis among men who have sex with men (MSM) has been reported from several U.S. cities, including Los Angeles, Seattle, Chicago, Miami, and New York City, possibly reflecting an increase in risk behavior in this population associated with increased wellness and well-being afforded by the availability of new, highly-effective antiretroviral therapy for HIV infection. From 1998 to 2001, outbreaks of early syphilis (including P&S and early latent) have been reported from these cities.

Appendix D Data Verification and Validation

The outbreaks in these five cities have been characterized by high rates of HIV co-infection. Although the total number of cases identified so far among MSM is relatively small, these outbreaks present a new challenge to attaining the national syphilis elimination objective of reducing the number of reported P&S syphilis cases to fewer than one thousand. Syphilis remains one of the most glaring examples of racial disparities in health, with 2001 rates among African Americans 16 times those among white Americans, down from a 64-fold differential at the beginning of the last decade. This racial disparity (16:1) is extreme compared to most other health outcomes including AIDS (9:1), infant mortality (2.5:1), and deaths attributable to heart disease (1.5:1).

Rates for Hispanics increased by 31.2% from 1997 to 2001. Communities burdened by poverty, racism, unemployment, low rates of health insurance, and inadequate access to health care are often disproportionately affected by syphilis. CDC aims to continue reducing this racial disparity in 2004.

Reduce the incidence of congenital syphilis: The lack of syphilis serologic testing and treatment during pregnancy remains the major reason that congenital syphilis persists in the U.S. Each positive test in a child is considered a medical emergency with immediate health services follow-up. The absence of testing is often related to complete lack of, or late initiation of, prenatal care. In 2001, 441 cases of congenital syphilis were reported to CDC, a rate of 11.1 cases per 100,000 live births. Now below the 2001 target of 12/100,000, this rate reflects a 59% decline in the number of cases since 1997 (1078 to 441 cases).

Tuberculosis (all):

Information on the percentage of TB patients reported in 2004 who complete TB treatment within 12 months will be available in June 2006. The last TB cases reported on December 31, 2004 will not have their 12-month treatment period completed until December 31, 2005. Then, 6-9 months are needed to tabulate, complete, verify, and report the data. This information is obtained from the national TB Surveillance System.

Information on the percentage of TB cases reported in 2004 with initial positive cultures and drug susceptibility results will be available by June 2005. This information is obtained from the national TB Surveillance System.

CDC recently revised the national reports for the data that addressed the following two measures: (1) Increase the percentage of contacts of infectious cases who are placed on treatment for latent TB infection and complete a treatment regimen; and (2) Increase the percentage of other high-risk infected persons who are placed on treatment for latent TB infection and complete a treatment regimen. For the first measure, the definition for contacts changed from contacts of "infectious cases" to "sputum smear-positive cases". The new system came on-line in CY 2000; the data for 1999 will not be representative because of the transition that occurred. The data for 2000 will not be submitted by the states until August 2002. Because the methods and definitions of reporting are substantially revised in the new system, data analysis will not yield results for these measures until after August 2003. Because of the change in definitions, program performance will appear to drop between 1998 to 1999, but the data are not comparable.

Information on the completion of treatment for latent TB infection for contacts of smear-positive cases who are started on treatment in 2004 will be available in mid-2006. Depending on the regimen used, it takes 2-9

Appendix D Data Verification and Validation

months to complete treatment. Therefore, some patients will not complete treatment until December 31, 2005. Approximately 6-9 months are allowed to tabulate, complete, verify, and report the data. This information is obtained from the national Aggregate Reports for TB Program Evaluation.

Information on the percentage of complete reporting of surveillance data items for TB cases reported in 2004 will be available by June 2005. This information is obtained from the national TB Surveillance System.

TB morbidity data and related information submitted via the national TB Surveillance System are entered locally or at the state level into CDC-developed software. The software contains numerous data validation checks. Data received at CDC are reviewed to confirm their integrity and evaluate completeness. Routine data quality reports are generated to assess data completeness and identify inconsistencies. These reports are shared with the reporting areas and discussed during site visits.

Data submitted via the national Aggregate Reports for TB Program Evaluation are checked for accuracy and inconsistencies. Problems are resolved by CDC staff working with state and local TB program staff. During regular visits to state, local, and territorial health departments, CDC staff review TB registers and other records and data systems and compare records for verification and accuracy. At the end of each year, data are again reviewed before data and counts are finalized and published.

Immunization

Data is obtained from a variety of sources, including the National Notifiable Disease Surveillance System (NNDSS), CDC, EPO; the National Congenital Rubella Syndrome Registry (NCRSR), CDC, NIP; the Active Bacterial Core Surveillance (ABCs), Emerging Infections Programs, CDC, NCID; and the National Health Interview Survey (NHIS), CDC, NCHS.

Public Health Improvement

REACH: Grantees will report on the development of implementation and evaluation plans, which will be reviewed by CDC staff. Site visits and data acquired by the CDC grant reporting system are also used. No data lags are expected.

The measure will be verified by the CDC grant reporting system.