Chapter 3

Collecting and Organizing Data

Contents

Chapter 3
Collecting and Organizing Data

Introduction

Throughout PATCH, data are used to guide the process. Phase II begins when the community group forms working groups to assess community health status and needs. The community group uses data as it selects health priorities in phases II and III. In phases III and IV, the community group may want to obtain additional information on selected health priorities, target groups, and related community resources as it designs appropriate and effective interventions. In phase V and throughout the process, data are collected and examined to evaluate interventions and identify areas for program improvement.

PATCH communities have found data valuable to

- serve as a motivator, helping to develop within the community a sense of ownership; communities are more supportive of interventions when decisions are based on data.

- substantiate information presented to the community about health issues.

- greatly increase the ability of the community to form partnerships and to compete for grants, funding from foundations, and other resources.

The data obtained should be those that the community deems relevant to its PATCH activities. The community should use existing data when available and collect new information when necessary. All communities should compare their data with data for the state or nation; some communities may also want to compare their data with data for a region, district, or neighboring county.

The community should use existing data when available and collect new information when necessary.

The highest quality data are those that are precise enough for scientific validity and are representative of the community. Although the community should strive to use the most valid and reliable data, sources of data or information that are not representative of the community can prove valuable for use in health planning. Using multiple data sources may help compensate for lack of more precise data and can provide an overview of the community that is valid for program development.

In phase II of PATCH, participants need to obtain and analyze quantitative and qualitative data that help reveal the leading health problems in the community. The quantitative data include vital statistics, disease registry, and survey data, and the qualitative data derive from opinion surveys of community leaders and, as appropriate, the community at large. The community may identify other sources of data as well.

During this phase of PATCH you, as local coordinator, must work with community members to

• obtain and analyze mortality, morbidity, and behavioral data.

• obtain and analyze community opinion information.

• present and use data to identify health priorities.

• determine ways to share data with the community.

This chapter addresses data issues related to community needs assessment and identification of health priorities. It describes the types of data the community should collect; methods for compiling data; and ways to analyze, organize, and display data. It also includes some conceptual information that may help you guide participants in understanding the importance of using data, how to document and describe the community's health problems, and how to compare community data with data for other areas. For additional assistance related to data, identify expertise within your community, consult the Bibliography (see Appendix 5), or consult with your state health department or a college or university.

Causes of death and disability

In the year 1900, most people died from infectious diseases. Gradually, as living conditions improved and advances were made in sanitation, immunization, and medicine, deaths from infectious diseases decreased. Chronic diseases then became the leading causes of death. Here are the leading causes of death in 1900 and 1990; the causes of death with the highest rates are listed first.

Leading Causes of Death, United States, 1900 and 1990	
1900	**1990**
Tuberculosis	Coronary heart disease
Pneumonia/influenza	Cancer
Diarrheal diseases	Stroke
Coronary heart disease	Injuries
Liver disease	Bronchitis/emphysema
Injuries	Pneumonia/influenza
Stroke	Diabetes
Cancer	Suicide
Bronchitis	Chronic liver disease
Diphtheria	Homicide

Accordingly, health officials have begun to turn their attention to preventing behavior-related diseases and to encouraging behavior and lifestyle changes at an individual and community level that can significantly influence whether a person develops a disease. Increasing attention has also been paid to intentional and unintentional injuries as causes of death. Elements of these events are believed to be controllable; many injuries can be prevented by changing related behaviors, policies, and developing a supportive environment.

Preventing premature death and disability

To prevent premature death and disability, we begin by asking why they occurred. Similarly, we look for "whys," or risk factors, for diseases. Factors that contribute to premature death include

- unhealthy behaviors.
- biologic or genetic conditions that predispose people to disease.
- inadequacies in the health care system.
- unsafe environmental conditions.

Unhealthy behaviors contribute the most to premature death and account for an estimated 54% of all diseases of the heart, 37% of cancers, and 69% of motor vehicle fatalities. In the United States, it is estimated that about 50% of all deaths before age 75 are caused by unhealthy behaviors, 20% by environmental factors, 20% by biologic factors, and 10% by inadequacies in the health care system.[1]

[1] Dever GEA. Community Health Analysis: Global Awareness at the Local Level (2nd ed.). Gaithersburg, MD: Aspen Publications Inc., 1991;35.

Risk factors can be either behavioral or nonbehavioral. Some risk factors–particularly those that are behavioral–may be amenable to change. Although we cannot change people's genetic makeup, we may be able to encourage them to engage in regular physical activity, eat healthy foods, take blood pressure medication, or stop smoking, and these behaviors may reduce the risk for certain diseases.

Environmental factors may also be changed to lower the risk for death or disability. For example, traffic lights might be installed at an intersection that is the site of many traffic fatalities or grocery stores might provide more low-fat food choices. Policy changes might also be possible–for example, laws prohibiting the sale of tobacco products to minors might be rigorously enforced, clean indoor air policies might be used to ensure smoke-free environments, and a worksite policy might allow flexible work schedules to encourage employees to engage in regular physical activity.

PATCH can be used to guide the community in analyzing causes of death and disability, identifying risk factors for these causes, and developing ways to reduce or change these risk factors, thereby reducing death and disability.

The following table, Table 1, summarizes risk factors that contribute to the leading causes of death.

Table 1. Contributors to the Leading Causes of Death

		Heart disease	Cancer	Stroke	Injuries (Nonvehicular)	Influenza/ Pneumonia	Injuries (Vehicular)	Diabetes	Cirrhosis	Suicide	Homicide
Behavioral risk factor	Tobacco use	•	•	•	•	•					
	Diet	•	•	P				•			
	Obesity	•	•					•			
	Lack of exercise	•	•	•				•			
	High blood pressure	•		•							
	High blood cholesterol	•		P							
	Stress	P		P	•		•			•	•
	Alcohol abuse	•	•	•	•		•			•	•
	Drug misuse	P	•	P	•		•			•	•
	Not using seatbelts						•				
	Handgun possession				•					•	•
Nonbehavioral risk factor	Biological factors	•	•	•		•		•	•	•	P
	Radiation		•								
	Workplace hazards		•		•		•				
	Environmental factors		•		•						
	Infectious agents	P	•			•			•		
	Auto/road design				•						
	Speed limits						•				
	Health care access	•	•	•	•	•	•	•	•	•	•

P = possible

Quantitative data

By analyzing mortality, morbidity, and behavioral data, we attempt to quantify health problems and the behavioral risk factors that contribute to them.

Mortality and morbidity data

Mortality data describe the deaths that occur in your community. These data are often analyzed by sex, race, age at death, and other variables about the deceased and compared with similar local, state, and national data. Such comparisons may indicate that certain health problems are more extensive in your community than in others or more severe among some groups than others.

Mortality data are collected in a standard format by all 50 states, the District of Columbia, Puerto Rico, and U.S. territories. The data are generally maintained by the state health department vital statistics units, from which they are available in computerized form for at least the last 10 years. These data are submitted to CDC's National Center for Health Statistics (NCHS) where they are combined with data from all the other states and territories in a national mortality data set.

For more information about WONDER, contact your state or local health department or CDC/Atlanta at (404) 332-4569.

Compiling information on all deaths that occur in a particular year and preparing a data set for public use takes a considerable amount of time. Consequently, complete files are usually not available from your state health department for several months after the end of a calendar year. Before then, preliminary files or advance data may be made available to health departments and other interested parties. For example, NCHS provides provisional mortality data in its Monthly Vital Statistics Report, which is available in many libraries, at most state vital records units, or through subscription. Summary mortality data are also available by state on CDC's computer network through WONDER (Wide-Ranging Online Data for Epidemiologic Research), which is designed by CDC to simplify online access data bases maintained by CDC. WONDER can be accessed from a modem-equipped personal computer.

Throughout the United States and most of the world, causes of death are standardized by translating the written diagnosis into a medical code contained in the *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, Ninth Revision* (ICD-9). To allow for comparison between community, state, and national data, the cause of death codes used by NCHS are taken from the ICD-9.

Mortality data are reported by place of residence or place of occurrence. The resident death rate reports the death rate for residents of an area, no matter where the deaths took place. The occurrence death rate reports the death rate for an area, regardless of where the decedents resided. The resident death rate is generally used for assessing the health risk of a community.

Morbidity data can be used to quantify the burden of disease in a population. These data help identify the leading causes of illness, injury, and disability and are useful in considering such outcomes as quality of life or the social and economic costs to society of certain health problems. Morbidity can be defined by looking at the incidence or prevalence of disease in the population. The incidence is the estimate measure of new cases of a disease or other events during a period of time. The prevalence is a measure of all cases, old and new, of disease or other health events that exist at a given time. The morbidity information provided by hospital discharge data and other data sources can assist your community in identifying the leading causes of disease and economic burden.

Although morbidity data are not always easily accessible and accurate, you are encouraged to look for existing sources. Morbidity data can be derived from medical records and population-based data sources. As with mortality data, these data can also be analyzed by subgroup or cause of morbidity and compared with data for other areas. The two main difficulties associated with collecting morbidity data are the following:

- Data are not reported in a consistent manner because there are no federal regulations for reporting morbidity data, reporting requirements vary between states and prevent comparability, and not all providers routinely maintain such data.

- In the case of hospital discharge data, the population served by local hospitals may not be the same as the community's population; local hospitals may serve people from outside the community; and residents may go outside the community for health care.

Hospital discharge data, which are sometimes available through a state agency or from local hospitals, are a major source of morbidity data. These data, which generally include reason for hospitalization and length of stay, can contribute to measuring the burden and cost of illness and disability in the community. Statewide hospital discharge databases, currently maintained by 35 states, generally support health planning and cost containment. Many states also maintain registries for such health priorities as cancer and birth defects. Your state health department should know if these databases exist and where they are located within your state.

Measures of mortality and morbidity

Expressing deaths as rates allows for useful comparisons.

Sometimes a count of health events (number of deaths or hospitalizations) is inadequate for expressing the burden of mortality and morbidity. The number of health events has little meaning unless the size of the population is known, for a higher count is expected from a larger population. Knowing that a county had 248 deaths due to cardiovascular disease and that the state had 7,658 deaths does not allow for comparison. Expressing these deaths as rates, 231.2 deaths per 100,000 for the county and 213.4 deaths per 100,000 for the state, does allow for useful comparisons. Values for different areas, populations, or periods must be made comparable to allow for evaluation. Thus mortality and morbidity data are usually presented as rates.

Crude, specific, and adjusted rates

A rate is expressed as a fraction in which the numerator is the number of persons to whom an event occurred during a period of time, and the denominator is the total number of persons in the population at risk for the event during the same period of time. The fraction is usually given per 1,000 or 100,000 persons per year.

A crude death rate is the number of deaths for a given population during a given period, divided by the total population during the same period. The population can be that of a state, a county, or another geographic unit for which data are available. The crude death rate is a rough estimate because it reflects all the factors in a community that could affect the total death rate, such as the demographic, social, and economic makeup of the community. Thus, crude rates can be misleading if an event occurs disproportionately in a group that is overrepresented or underrepresented in the population. For example, a high crude death rate is generally expected for a community with a high proportion of older persons, such as a retirement community.

To more clearly identify priorities and target populations, your community should calculate rates for specific subgroups, by age, sex, race, or other characteristics. For example, you might calculate an age-specific rate for persons aged 18 to 24 years or an age-sex-race-specific rate for black women aged 18 to 24 years. For these rates, the number of events for the subgroup is divided by the total population of the subgroup.

To focus on specific diseases, you can calculate cause-specific rates for particular causes of death or disability. For example, you might calculate the mortality rate for all cardiovascular deaths in your community. You might also calculate cause-age-sex-race-

specific rates for subgroups in your community, for example, 45-to-64-year-old white men who die of cardiovascular disease.

An adjusted rate, or a standardized rate, helps control for demographic differences between populations being compared. Adjusted mortality rates, for example, estimate what the mortality rates for populations would be if their composition were similar to that of a comparison, or standard, population. The standard population could be the state or U.S. population, and adjustment can be made for age, race, sex, or other characteristics. Because the rates are standardized to the same population, however, they are comparable and thus allow for analyses between populations.

Choosing a rate

Whether to use crude, specific, or adjusted rates depends on the purpose of the analysis. Crude rates are easy to compute and understand. They represent the actual experience of the population and provide a rough measure that reflects many factors, such as the demographic, social, and economic makeup of the community. Crude rates are useful when determining the overall allocation of health resources. For public health planning, however, specific rates provide the most detailed information about the pattern of disease in a population. However, comparisons of specific rates can become cumbersome when data for many groups must be presented. Adjusted rates remove the effect of population differences to allow for comparisons between groups or over time. Bear in mind that the actual value of an adjusted rate has no inherent meaning because it is mathematically derived and can be based on one of many standard populations. Therefore, adjusted rates should be used only for comparison purposes.

For public health planning, specific rates provide the most detailed information about the pattern of disease in a population.

You may find that a combination of measures is needed to analyze the data collected to fully express the disease burden in your community. Also, examining the same rate over time, or trend analysis, would provide evidence of the increasing, decreasing, or stable trends in mortality or morbidity.

Years of potential life lost

Another way to examine the impact of disease is in years of potential life lost (YPLL) for each death that occurs before a selected age end point. Therefore, YPLL is a measure of premature mortality. The generally accepted end point is age 65; however, because the average life span in the United States is longer than in many countries, premature mortality is often defined nationally as deaths before age 75 or even 85. Alternatively, the end point can be based

YPLL is a measure of premature mortality.

on the life expectancy of persons in specific age groups. YPLL then is the difference between the age at death and the selected end point.

YPLL provides another way to assess the burden of deaths due to specific diseases. Because YPLL is greater for deaths among young persons than old persons, it emphasizes the causes of death that affect the young. It can be useful for expressing the impact of premature deaths from injuries or from chronic diseases, such as breast cancer and coronary heart disease. Calculating the number of years of life lost due to selected causes of death will help the community view these causes of death not only as life lost to individuals but to the community.

Additional measures for determining the impact of preventable disease, death, disability, and injury are currently being refined and may become available in the future. These include measures that indicate "quality of life" and "years of healthy life." Contact your state health department concerning progress made with additional measures.

Collecting mortality and morbidity data

As you proceed with your working group and with assistance from resources such as your state health department or university, obtain the data required to complete the appropriate pages in the Program Documentation (PD) tool (see Appendix 3). The sections to be completed while compiling mortality data are

- Unique Health Events (PD-II)

- Number of Deaths and Years of Potential Life Lost by Major Disease Categories (PD-III)

- Five Leading Causes of Death by Age Groups (PD-IV)

- Comparison of Mortality Rates for Leading Causes of Death by Race, Sex, and Age Groups (PD-V)

A unique health event is an event or activity that may have a short- or long-term effect on the health or health risk of a community. A devastating hurricane or passage of a seatbelt law might explain abrupt variations in mortality data that should be discussed with the community group. Provide trend data on at least the top five causes of death. When analyzing data for communities with small populations, it might be appropriate to use three to five years of aggregated data. Otherwise the community group might identify health priorities that reflect one-time events.

As you coordinate data working groups, consult with the working group collecting opinion data to find out what key problems have been identified. As appropriate, collect additional mortality data concerning these problems.

The disease categories in PD-III are based on the leading causes of death. Decide if these, or other, categories are appropriate for your community. When obtaining the number of deaths by disease categories and age group (PD-III and V), be consistent with how you define the cause of death when comparing local data with state and national data. The ICD-9 codes are used frequently by CDC to define the leading causes of death.

Frequently Used ICD-9 Codes

Coronary heart disease (Heart disease)	ICD codes 390-398, 402, 404-429
Cancer	ICD codes 140-208
Stroke	ICD codes 430-438
COPD*	ICD codes 490-496
Unintentional injuries†	ICD codes E800-E949

When making a comparison of mortality rates by race, sex, and age groups (PD-V), you will need to aggregate three to five years of data unless your community is a large metropolitan area. If your community has data on additional populations within the community, such as Hispanics, Asian-Americans, and Native Americans, prepare similar PD pages to record those data.

The morbidity data a community is able to collect vary. Therefore, there are no forms in the PD for displaying morbidity data. Communities can generally obtain hospital discharge data and display them for their community group. When you are obtaining or presenting hospitalization data, you should distinguish what discharge diagnosis is used. Most forms of hospitalization data may be tabulated by primary discharge diagnosis, any listed discharge diagnosis, or both. Reporting only primary diagnosis will often underrepresent the actual prevalence of a particular disease condition, whereas reporting any listed diagnosis may more accurately represent the disease burden associated with a disease condition.

Chronic obstructive pulmonary disease (COPD) includes such categories as bronchitis and emphysema.

†*Unintentional injuries include such categories as falls, drownings, fires and burns, poisonings, and motor vehicle injuries.*

As you coordinate data working groups, consult with the working group collecting opinion data to find out key priorities identified. As appropriate, be able to provide additional morbidity data concerning major priorities.

Explore the availability of mortality and morbidity data within your community and state and from your various partners. Some of the resources for community data include

- State and local health departments–Census data on births, deaths, and social conditions such as divorces.

- State and local social service departments–Variety of data including percentages of population on welfare and unemployment.

- State and local departments of highway safety–Data on traffic injuries and seatbelt use.

- State and local police departments–Information on crime trends, high-crime areas, and driving under the influence of alcohol or drugs.

- Boards of education–Information on percentage of population who have secondary degrees and undergraduate degrees (also in census data).

- Voluntary agencies–Data on causes of death, disability, and their risk factors.

- Hospitals–Information on length of stays, major causes of hospitalization, and description of hospitalized populations.

- Major employers or chamber of commerce–Data on demographic information on workforce, illness, and disability information.

- Colleges and universities–Information on morbidity trends and forecasts on population trends.

The following pages are sections of the Program Documentation (see Appendix 3) that relate to mortality data. You will need to create additional pages for each age group listed in PD-V. The age groups listed in the year 2000 national health objectives should be used: <1, 1-14, 15-24, 25-44, 45-64, and 65+. Data should be prepared for at least "whites" and "blacks and others." If you have sufficient data, prepare similar pages on additional populations such as Hispanics, Asian-Americans, and Native Americans.

PD-II. Unique Health Events

A unique health event is an event or activity that takes place in the community that may have a short-term or long-term effect on the health or health risks of its citizens.

Examples of unique health events include special community health promotion and health education activities, health legislation, and environmental or natural events. Events can have a negative or positive effect on health. For instance, positive events might include the addition of fluoride to the drinking water or passing a law requiring the use of seatbelts. Negative events might include a hurricane or flood or the repeal of the tax on tobacco products.

Do not report PATCH program results or activities in this section. Report PATCH activities in PD-XII.

Date	Description of the Event	Number of People Affected

PD-III. Number of Deaths and Years of Potential Life Lost
by Major Disease Categories*

Community _____ Year _____

Disease Category	Rank	# of Deaths	YPLL† Before Age 75
Heart disease			
All cancers			
Lung cancer		‡	‡
Cerebrovascular disease			
Emphysema			
Influenza and pneumonia			
All fatal injuries			
Motor vehicle injuries		‡	‡
Liver disease			
Suicide			
Homicide			
Diabetes mellitus			
Other			
Total			

*   Based on leading causes of death.

†   YPLL = Years of potential life lost for deaths > 1 year of age.

‡   To calculate the "Total" number of deaths or YPLL, add all numbers in
       the column except for lung cancer and motor vehicle injuries.

Source: _____

PD-IV. Five Leading Causes of Death by Age Groups

Community _____ Year _____

Age < 1 Cause	Total*	Percent†
1. _____		
2. _____		
3. _____		
4. _____		
5. _____		
All other causes		

Age 1-14 Cause	Total*	Percent†
1. _____		
2. _____		
3. _____		
4. _____		
5. _____		
All other causes		

Age 15-24 Cause	Total*	Percent†
1. _____		
2. _____		
3. _____		
4. _____		
5. _____		
All other causes		

* Total number of deaths from this cause.
† Number of deaths from this cause divided by total of all deaths in this age group.
Source: _____

PD-IV. Five Leading Causes of Death, by Age Groups in

Community_____Year_____

Age 25-44 Cause	Total*	Percent†
1. _____		
2. _____		
3. _____		
4. _____		
5. _____		
All other causes		

Age 45-64 Cause	Total*	Percent†
1. _____		
2. _____		
3. _____		
4. _____		
5. _____		
All other causes		

Age 65+ Cause	Total*	Percent†
1. _____		
2. _____		
3. _____		
4. _____		
5. _____		
All other causes		

* Total number of deaths from this cause.
† Number of deaths from this cause divided by total of all deaths in this age group.
Source: _____

PD-V. Comparison of Mortality Rates for Leading Causes of Death by Race, Sex, and Age Groups

Mortality Rates (per 100,000) for Leading Causes of Death Among Males Aged ____

Community _____ / (Year)   State _____ / (Year)   United States ___ / (Year)

Rank	Cause	# of deaths	Rate	Rank in state	# of deaths	Rate	Rank in nation	# of deaths	Rate
1.									
2.									
3.									
4.									
5.									

Mortality Rates (per 100,000) for Leading Causes of Death Among Females Aged ____

Community _____ / (Year)   State _____ / (Year)   United States ___ / (Year)

Rank	Cause	# of deaths	Rate	Rank in state	# of deaths	Rate	Rank in nation	# of deaths	Rate
1.									
2.									
3.									
4.									
5.									

PD-V. Comparison of Mortality Rates for Leading Causes of Death by Race, Sex, and Age Groups

Mortality Rates (per 100,000) for Leading Causes of Death Among White Males Aged ____

Community _____ / _____ State _____ / _____ United States / _____
(Year) (Year) (Year)

Rank	Cause	# of deaths	Rate	Rank in state	# of deaths	Rate	Rank in nation	# of deaths	Rate
1.									
2.									
3.									
4.									
5.									

Mortality Rates (per 100,000) for Leading Causes of Death Among White Females Aged ____

Community _____ / _____ State _____ / _____ United States / _____
(Year) (Year) (Year)

Rank	Cause	# of deaths	Rate	Rank in state	# of deaths	Rate	Rank in nation	# of deaths	Rate
1.									
2.									
3.									
4.									
5.									

PD-V. Comparison of Mortality Rates for Leading Causes of Death by Race, Sex, and Age Groups

Mortality Rates (per 100,000) for Leading Causes of Death Among Black and Other Males Aged ____

Community ____ / ____ State ____ / ____ United States ____ / ____
(Year) (Year) (Year)

Rank	Cause	# of deaths	Rate	Rank in state	# of deaths	Rate	Rank in nation	# of deaths	Rate
1.									
2.									
3.									
4.									
5.									

Mortality Rates (per 100,000) for Leading Causes of Death Among Black and Other Females Aged ____

Community ____ / ____ State ____ / ____ United States ____ / ____
(Year) (Year) (Year)

Rank	Cause	# of deaths	Rate	Rank in state	# of deaths	Rate	Rank in nation	# of deaths	Rate
1.									
2.									
3.									
4.									
5.									

Preparing for community group meeting

As you and the working group analyze and organize the data for presentation to the community group, prepare to give an overview of the items listed in the Program Documentation and then spend time on the causes of death that are especially noteworthy. In particular, emphasize those causes of death that have a higher rate when compared with the state, nation, or other populations and those that are of special concern to the participants in the community group or respondents to the opinion survey. If the numbers in the data set are small, consult with a statistician or epidemiologist on how to interpret them. Three to five years of data may need to be combined to have numbers large enough to be of value in health planning. Also, looking at trends in the data for at least the top five causes of death might prove valuable.

Prepare a Mortality Data Packet containing a matching set of overheads and handouts.

Prepare a mortality data packet containing a matching set of overheads and handouts to display your community's data. Examples of overheads showing community data are located as handouts for the phase I meeting. Review the section on *Presenting Data* in this chapter. The packet might include completed pages copied from the Program Documentation. It should display summary information and highlights from the Program Documentation and should address at a minimum the following items:

• unique health events that might affect interpretation of the community's data (PD-II)

• five leading causes of death in the community (PD-III), compared with state and national data if possible

• five leading causes of death broken down in categories that show substantial differences: by race, sex, or age (PD-IV and V)

• display the five leading causes of YPLL (PD-III)

• trend data for five leading causes of death

• any comparison or group (race, sex, age) data (PD-V) that are outstanding or unique

Prepare a morbidity data packet containing a matching set of overheads and handouts to display your community's data. The packet should concentrate on the top five reasons for being hospitalized and any other data the group can identify.

Behavioral data

Because behavioral risk factors contribute to almost half of the leading causes of death and disability and have the potential to be changed, behavioral data are an important part of the PATCH process. Behavioral data should be used by the community to identify risk factors and target populations to address. Thus you will need to determine the prevalence or frequency of behavioral risk factors in your community. There are two approaches to obtaining these data: use existing data to estimate the prevalence of particular risk factors in your community or collect the information from the community.

Using existing behavioral data, including the BRFSS

All 50 state health departments have statewide behavioral data collected by using CDC's Behavioral Risk Factor Surveillance System (BRFSS). These data estimate the prevalence of behaviors related to the leading causes of death and disability and are available by such variables as age, race, sex, income, and education. In 1981, CDC began helping states and communities to survey adults, 18 years and older, by telephone about their health behaviors. In 1984, CDC initiated BRFSS to enable states to collect state-specific behavioral risk factor data. By 1994, all 50 states and the District of Columbia were participating in the BRFSS.

By 1994, all 50 states and the District of Columbia were participating in the BRFSS.

BRFSS is a telephone survey for which participating states use a standard protocol and standard interviewing methods to complete from 100 to more than 300 interviews each month. The interviewing instrument used for the BRFSS is in three parts: the core questionnaire, standard modules, and state-specific questions. The core questionnaire is a set of questions asked by all states; it deals primarily with recent or current behaviors that are risk factors for disease or injury and with quality of life indicators. Responses to certain core questions are used to define the risk factor variables that are used to classify respondents as either at risk or not at risk for adverse health events. (See Risk Factor Definitions provided with PD-VII in this chapter.) Each year, CDC supplies standard reports on the core data to states. These reports include the prevalence of risk factor variables by demographic characteristics and among participating states.

The standard modules are sets of questions developed by CDC on specific topics suggested by states. States decide which, if any, standard modules they will include each year. In addition, states develop and use state-specific questions. Over the years, topics covered by the standard modules and by state-specific questions

have included screening for cervical, breast, and colorectal cancer; health care visits; health insurance; smoke detector use; smoking cessation; perceived health problems; and sources of health information. For further information on questions, ask your state coordinator.

As you explore other sources of behavioral data within your community and state, look for data that also address risk behaviors in other populations not addressed in the BRFSS. For example, in 1993 the CDC's Youth Risk Behavior Survey was used by 43 states and territories and 13 large cities to determine behaviors among ninth to twelfth grade students. For more information, contact the health education coordinator in your state education department or the CDC, Division of Adolescent and School Health at 770/488-5330.

Community use of BRFSS data

Most state health departments can provide communities with state and national behavioral data. Many states also combine multiple years of BRFSS data to have a sample large enough to provide data for subsets of the state. For example, one state has four public health regions and provides BRFSS data for each region. Another state determined it was more appropriate for health planning to analyze its data for three subsets of the state: the one large metropolitan area, all other urban counties, and all rural counties.

Most states may find that sample sizes are inadequate for estimating risk factor prevalence for small geographic areas such as counties or your PATCH community. However, *synthetic estimates* of community data can be derived by extrapolating from the BRFSS data for the state or subset of the state. For example, you can calculate synthetic estimates by using the population estimates for the subgroup of interest and the state BRFSS risk factor prevalences for that subgroup. This approach assumes that the risk factor prevalences for specific subgroups in your community are the same as the statewide risk factor prevalences for the same subgroups. For example, it assumes that black women in your community have the same prevalence of smoking as black women throughout the state. The accuracy of the estimate depends on whether the residents in your community behave similarly to residents in the state as a whole.

Synthetic estimates derived from state and national sources are crude and can be misleading, particularly when individual communities differ demographically from the state as a whole. However, when used in conjunction with other data, synthetic estimates may be sufficient for health planning, for setting priorities, and for

targeting prevention strategies. In addition to BRFSS data, you can also use other national data sets, such as the National Health Interview Survey, to derive the synthetic estimates for communities.

Another caution in using synthetic estimates concerns dividing the state BRFSS sample by more than one demographic variable. Too many demographic divisions of the sample make the size of the individual cells too small for accurate data analysis. Making projections about your community on the basis of such small numbers may be inappropriate. A rule of thumb may be that fewer than 50 in a cell restricts the interpretation of these data. Synthetic estimates remain a quick and inexpensive way to use state, regional, or national data to estimate the prevalence of risk behaviors in your community.

Collecting behavioral information

If you do not wish to use state BRFSS data or synthetic estimates for your community or subset of the state, you can obtain behavioral data for your community. Some state health departments have used the states' BRFSS mechanisms to conduct a special survey of a PATCH community or to oversample in the community while performing their monthly BRFSS calls. Other PATCH communities have conducted their own behavioral survey with assistance from their state health department. If you wish to do a community survey, talk with your state coordinator about whether the state can provide assistance or whether there are other available resources, such as university polling centers.

If you choose to collect your own behavioral data, you may want to add additional questions about other risk factors. These questions will need to be reviewed by your state coordinator or another expert in questionnaire design to ensure they will provide the information you want. Because the established computer software cannot be used to analyze added questions, you will need to analyze separately any additions to the standard questionnaire.

A community behavioral survey may cost from $3,000 to $10,000. The cost primarily depends on whether interviewers are paid or are volunteers. Paying interviewers is strongly recommended. The collecting of behavioral data requires planning, training interviewers, and coordinating resources.

Conducting a community behavioral survey has advantages:

• The data are specific to the community and provide valuable information about its citizens.

- The behavioral survey is a major activity that can help mobilize the community early in the process. It also gives visibility to PATCH because many community members hear about PATCH as a result of the survey.

Disadvantages include

- Depending on whether you pay to have the survey done or rely on volunteers, the survey may use a significant amount of limited human and financial resources.

- The time it takes to collect and analyze the data can cause the community to lose interest.

Your community may wish to collect behavioral risk factor information by developing its own survey instrument and protocol. This may prove to be an even more difficult and time-consuming process. Before you proceed, contact your state health department or another resource for assistance with instrument and survey design to ensure that the data are usable and are as comparable with other data sets as possible.

Making the choice

As you make decisions on what behavioral data to obtain, it would be advisable to work with the state coordinator and the person in charge of the BRFSS to examine existing data. Some PATCH communities have chosen simply to use BRFSS data for their state or subset of their state. Other communities believe the subgroups in their community are similar to the same subgroups within the state and accept synthetic estimates based on state data. Still other communities believe that members of their community are different from the general population of the state and that conducting the survey locally is worth the effort because of the unique information they would get.

The choice of what method to use to gather risk-behavior information is not always easy. Individual states and communities must look carefully at their alternatives and make a choice, based on their needs and resources. It is, however, a decision that needs to be made early in the process. Whatever the decision, as you proceed with your working groups and with assistance from resources such

PD-VII. A Comparison of Behavioral Data (Percentage) Among Adults by Community, State, and Nation

	Community			State			Nation		
	M	F	Total	M	F	Total	M	F	Total
Seatbelt (1)									
Seatbelt (2)									
Hypertension (1)									
Hypertension (2)									
Overweight (1)									
Overweight (2))									
Current smoking									
Acute (binge) drinking									
Chronic drinking									
Drinking and driving									
Sedentary lifestyle									
No leisure-time activity									
Regular and sustained activity									
Regular and vigorous activity									
Cholesterol screening (1)									
Cholesterol screening (2)									
Cholesterol awareness									

Community data source: —————————————————

State data source: —————————————————

National data source: —————————————————

as your state health department or university, obtain the data needed to complete the PD-VII form in the Program Documentation.

Behavioral Risk Factor Surveillance System

Risk Factor Definitions

Seatbelt (1)	Respondents who report that they "sometimes," "seldom," or "never" use safety belts.
Seatbelt (2)	Respondents who report that they "nearly always," "seldom," or "never" use safety belts (i.e., do not always use a safety belt).
Hypertension (1)	Respondents who report that they have had their blood pressure checked within the past two years.
Hypertension (2)	Respondents who report that they have ever been told they are hypertensive.
Overweight (1)	Respondents who report that they are at or above 120% of ideal weight. Ideal weight defined as the midvalue of a medium frame person, from the Metropolitan Life Insurance height-weight tables (1959).
Overweight (2)	Women with body mass index (weight in kilograms divided by height in meters squared (w/h**²) $>=27.3$ and men with body mass index $>=27.8$.
Current smoking	Current regular smoker (ever smoked 100 cigarettes and smoke regularly now).
Acute (binge) drinking	Respondents who report that they have had five or more drinks on an occasion, one or more times in the past month.
Chronic drinking	Respondents who report that they have had an average of 60 or more alcoholic drinks a month.
Drinking and driving	Respondents who report that they have driven after having had too much to drink, one or more times in the past month.
Sedentary lifestyle	Respondents who report that they have had no activity, or no physical activity, or pair of activities that were done for less than 20 minutes, or less than three times a week in the past month.

No leisure-time physical activity — Respondents who report that they have had no leisure-time physical activity during the past month. This measures Year 2000 Objective 1.5 - Target $\leq 15\%$.

Regular and sustained physical activity — Respondents who report that they have had physical activity 5 or more sessions per week, 30 minutes or more pre session, regardless of intensity. This measures Year 2000 Objective 1.3 - Target $\geq 30\%$.

Regular and vigorous physical activity — Respondents who report that they have had physical activity or a pair of activities for 3 or more sessions per week, 20 minutes or more per sessions, at 50% or more capacity. This measures Year 2000 Objective 1.4- Target $\geq 20\%$.

Cholesterol screening (1) — Respondents who report that they have ever had their blood cholesterol checked.

Cholesterol screening (2) — Respondents who report that they have had their blood cholesterol checked within the past five years.

Cholesterol awareness — Respondents who report that they have been told their cholesterol is high by health professional.

Preparing for community group meeting

As you analyze and organize your behavioral data, you will want to summarize the items listed in PD-VII and then spend time on behaviors that are especially noteworthy. Noteworthy behaviors include those with a higher prevalence in your community when compared with the state, nation, or those that have been expressed as a concern by the community group participants or the opinion survey respondents, or those that cause a high degree of death, disability, and illness.

Prepare a Behavioral Data Packet containing a matching set of overheads and handouts to display your community's data.

When you fill out PD-VII, a rule of thumb is to have at least 50 persons represented in any cell. If fewer than 50 persons responded, your sample size may be too small to quote but may be of some value in decision making. You can often correct this problem by combining variables such as age groups. Ask your state coordinator or statistician for help. When reporting the data, you should round off the prevalence figures to whole percents.

Prepare a Behavioral Data Packet containing a matching set of overheads and handouts to display your community's data. It might also include a copy of PD-VII. Review the section on *Presenting Data* in this chapter. The packet should display summary information of the behavioral data for community, state, and national levels and more detailed information when behavior varies by subgroups such as among men and women.

Qualitative information

By looking at opinion data from community leaders, we attempt to learn about the perceived health priorities and quality of life in the community. One of the key components of community-based health promotion is the active involvement of the community in the planning, implementation, and evaluation of health programs. The opinions or beliefs that come directly from community leaders can provide valuable information about the health needs of the community. We cannot emphasize enough the importance of this kind of information to the success of a community-based process. The opinions of the community must be heard and respected if there is to be community ownership.

Community opinion data

Opinion information provides viewpoints from the community about health awareness, needs, and perceived health problems. A combination of quantitative and opinion data will help the community define problems and develop meaningful community goals and objectives. A comparison of quantitative information (mortality, morbidity, and behavioral data) with the opinion data will either substantiate or disprove the opinions of the community.

For example:

One community perceived cancer to be a problem and requested assistance from its state health department. Members of the community believed that the drinking water was the culprit. As the community undertook the PATCH process, data were gathered and analyzed that showed lung cancer was the prominent type (site) of cancer, whereas digestive cancers were extremely rare. Thus, the community designed an intervention program to target lung cancer by reducing tobacco use.

Knowing what the community or specific populations perceive to be their health needs is extremely important in planning programs to address those needs. Opinion information also reveals the community's level of awareness of health issues and its health problems and can help direct the design of press releases and educational information. It reflects community values and other qualitative factors not provided by the quantitative data. Also, the process of collecting opinion data provides an opportunity to inform more community members about PATCH and helps build community support for planning and carrying out health programs. It also helps you identify sources of support and opposition within the community.

There are two types of surveys that can be done—one of community leaders and one of the community at large. The materials provided in this chapter emphasize the survey of community leaders. The Community Leader Opinion Survey questionnaire (see page CG3-29), working group materials, and a sample Communitywide Opinion Survey are included in the handouts.

Conducting a survey of the community at large can be a big task. To be representative, respondents should be randomly selected from the entire community. However, most community groups lack the time and resources to conduct this type of survey. Some communities collect communitywide data in malls, at health fairs, or by placing mail-in coupons in the newspaper. Although these data are not representative enough to weigh heavily in the decision-making process, they may be valuable by increasing the awareness of the community group and identifying issues that may need to be explored further when designing interventions later in the PATCH process. They should be analyzed separately from the leader survey.

Identifying opinion leaders

Every community has what it calls "opinion leaders." They are people in positions of power who have the reputation for getting things done, who made key decisions on previous issues, who actively volunteer their time to help the community, or who are formal or informal neighborhood or community leaders. Whereas some opinion leaders are easy to identify because they hold official positions in the community, there are many other leaders who are not in positions of authority, but are influential and knowledgeable about the community. Special efforts should be made to identify these informal leaders to obtain their opinions. To help community

groups identify key leaders to be interviewed, we have provided a list of some characteristics and affiliations on the Description of Respondent form page (see page CG3-30) with the Community Leader Opinion Survey materials. The final list of interviewees should represent the sex, age, race, and affiliation groups reflective of the demographic makeup of the community. Many PATCH communities generate 100 to 150 names and complete 60 to 125 surveys.

Community Leader Opinion Survey

1. What do you think the main health problems are in our community?

2. What do you think are the causes of these health problems?

3. How can these problems be reduced or eliminated in our community?

4. Which one of these problems do you consider to be the most important one in our community?

5. Can you suggest three other people with whom I might talk about the health problems in our community?

Thank you for your help. Right now I do not have any more questions, but may I contact you in the future if other issues come up.

Description of Respondent

Respondent's name_____

Record the following information for each respondent, without input from the respondent, if possible. To ensure confidentiality, separate this page from the rest of the survey before returning both to the working group chairperson or local coordinator.

1. Sex: ___Female   ___Male

2. Race: ___White   ___Black   ___Hispanic

   ___American Indian   __Asian   __Other

3. Age: _<18   _18-24   _25-44   _45-64   _65+

4. Affiliation that resulted in respondent being selected:

___ A. Business person

___ B. Citizen activist

___ C. City/county official

___ D. Civic association member

___ E. Community outreach worker

___ F. Health professional (specify)

___ G. Law enforcement person

___ H. Leader of organization of faith

___ I.   Local celebrity

___ J.   Media/news person

___ K. Neighborhood formal/informal leader

___ L. School board member/administrator/teacher

___ M. Social services provider

___ N. Voluntary health agency representative

___ O. Youth peer leader

___ P. Other

5. Member of community: _____<3 _____3-9 _____10+ years

6. Geographical area:_____ urban _____ rural

   Neighborhood: _____

Collecting opinion data

Communities have found it best to do the opinion survey of community leaders using person-to-person interviews. Preparing your community group to do the Community Leader Opinion Survey is a major component of the community group meeting for phase I (see Meeting Guide for Phase I). Then the survey is completed and the data are reported to the community group during phase II meetings.

There are seven steps to completing an opinion survey:

1. Identifying opinion leaders to be interviewed.

2. Developing additional questions, if desired, for the survey instrument.

3. Identifying and training interviewers.

4. Interviewing opinion leaders.

5. Collating data from the interviews and preparing handouts and overheads to present to the community group.

6. Completing the PD-VI Community Leader Opinion Survey Data.

7. Writing a final report based on the group's consensus of high-priority problems.

The community group may need to be involved in steps 1 and 2, but major responsibility for completing the remaining steps rests with the Opinion Data Working Group. The form to be completed while compiling opinion data is the Community Leader Opinion Survey Data.

The most practical way to analyze community opinion data is to simply rank the problems according to the frequency of their being stated as problems. To rank these data, review the information from interviews and tabulate the number of times a problem was stated. Group responses consistently and only when they are similar. If more than one person tabulates responses, have them work together to ensure that responses are grouped similarly.

A sample Community Leader Opinion Survey questionnaire packet is included in the handouts. The questionnaire and introductory scripts should be adapted as appropriate for your community. The packet includes these elements

• script for making the appointment

• introductory script

• sample questionnaire

• description of the respondent

• handouts and task sheets for working group

If you add questions to the survey, begin with easy-to-answer questions to help the respondent to feel comfortable. Ask questions in a logical order, covering one point completely before going on to another. Each interview should take no longer than 20 to 30 minutes. Keep in mind these additional guidelines when developing questions.

• Language problems can arise from ethnic, racial, or regional differences. Make sure that the respondent will understand the meaning of questions without much explaining.

• Avoid professional jargon. Rather than saying "we are collecting community opinion survey data in our community," you might want to say, "we are asking leaders their opinions regarding health problems in our community."

• Avoid asking two questions in one sentence. For example, "do you believe exercise breaks should be regularly scheduled in our schools?" This question should be divided: "Do you believe scheduled exercise breaks should be offered in our schools? If so, how often?"

• Avoid biased questions: "You think we need more health education in the schools, don't you?"

• Because lengthy questions are easily misinterpreted and tire both the respondent and the interviewer, brief questions are recommended. Words should be simple—questions should be short.

• To make responses to sensitive questions easier, provide a checklist of responses or a series of statements that the respondent can agree or disagree with, or use a vignette or story describing a situation and have the respondent comment about it.

PD-VI. Community Leader Opinion Survey Data

Data collection method

Number of interviewers

Total number
of people interviewed

From: _____ To: _____
Date collected

Rank	Health problem	Number of persons identifying problem	Percentage of persons identifying problem
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			

Source:

Preparing for community group meeting

As you and the working group analyze and organize the opinion data for presentation to the community group, prepare to give a listing of the responses as well as the top 10 items listed in the Program Documentation page. In particular, emphasize those items that have the most responses and those items that are substantiated by quantitative data.

Prepare a Community Opinion Data Packet containing handouts and overheads to display your community's data. The packet should list all responses from the Community Leader Opinion Survey and emphasize questions 1 and 4. Question 1 states, "What do you think the main health problems are in our community?" Question 4 states, "Which one of these problems do you consider to be the most important one in our community?" Record the top 10 responses for question 4 on the PD-VI page, Community Leader Opinion Survey Data. This page would then be included in the packet.

If you have data from your phase I exercise (see Meeting Guide) in which participants interviewed one another with the Community Leader Opinion Survey, you may want to collate the information as described previously and place it on a separate copy of PD-VI. If the group collects communitywide opinion data, rank responses by frequency and report in a format similar to PD-VI. Be certain you do not combine data from different sources. Review the section on *Presenting Data* in this chapter.

In addition to a ranking of the health problems, you will want to discuss additional insights such as the level of awareness in the community of major health problems as well as potential allies and the likelihood of their support.

As you organize your opinion data, note when problems identified through the quantitative data and through community opinion do not reflect the same concerns. You will need to determine whether more data, quantitative or opinion, should be collected to present all the information the community group will need to determine priority health problems.

Presenting data

As you present data to your PATCH community group, keep in mind the following principles (in appendixes 2 and 5, see the Tipsheets and the sections of Bibliography on Data Analyzing and Display and Epidemiology):

Prepare a Community Opinion Data Packet containing handouts and overheads.

1. *Present data in a simple, straightforward manner.* The more understandable data are to community group members, the more likely members will be to use the data in planning health interventions.

2. *Include a frame of reference for the data.* For instance, compare your community's data with national data, state data, or data from similar communities.

3. *Explain any limitations of the particular data set.*

4. *Be sure any presentation of data on paper can stand alone, regardless of the particular format you choose.* Label tables, charts, and graphs; specify the data source.

5. *Be sure that the measure you select for display (count, percentage, rate, or their measure) is appropriate for your message and the constraints of your graphic display.*

6. *Choose the graphic display most appropriate for your task.* When you present health data to help with health planning, you may want to use a graphic display to focus attention on differences between diseases, population groups, or other variables. Some graphic aids are more appropriate than others for illustrating certain types of data, fostering comparisons, and allowing your audience to quickly grasp important points. Which visual display you choose depends on the message you want to convey.

Here are a few basic guidelines for using the most popular charts and graphs:

• Select only one main message per visual display. If you want to focus on several aspects of the data, consider making a set of visual aids.

• A horizontal bar chart can be used to focus attention on how one category differs among several groups.

• A vertical bar chart is often most appropriate when you want to focus attention on a change in a variable over time.

• Consider using a cluster bar chart when you want to contrast one variable among multiple subgroups.

• A line graph can be used to plot data for several periods and show a trend over time.

• A pie chart is sometimes used to show the distribution of a set of events or a total quantity.

Adapting phase II to address a specific health issue or population

When a health issue of high priority is selected before beginning the PATCH process, some of the activities related to data collection are modified accordingly. For example, if the health problem to be addressed is cardiovascular disease (CVD), your community might want to modify the mortality pages of the Program Documentation and complete the appropriate sections with only CVD-related data. Or it might wish to collect some of the data recommended in the Program Documentation to determine the relative importance of CVD in your community. In most U.S. communities, CVD is a leading cause of both death and years of potential life lost. Similarly, the opinion survey might be modified to address CVD only or some of the general questions provided might be used with follow-up questions related to CVD. Behavioral data obtained might be limited to risk factors for CVD.

As you work with your community group, working group, and partners to identify data sources specific to CVD, the material in this chapter may prove valuable in helping you to collect, display, and present those data. When using PATCH to address the health needs of a specific population, such as older adults, make similar modifications.