1. Determining the Data You Need

Before beginning a search for data, it is important to outline and define the questions you need the data to answer. To determine the data you need, answer these questions:

- What do I need the data to do?
- What message does the data have to deliver?
- Who are the audiences for the data, and what data does each audience need?

What do you need the data to do?

- Define the problem
- Show that your solution alleviates or solves the problem
- Show the negative consequences of not using your solution
- Measure program outcomes

At any given time, you may need the data to do all these things, some of these things, or just one. To determine what you need the data to do, get a picture of your social, economic, and political environment:

1. What are the key public health issues for your community?

2. Define the problem central to the issue: i.e. access to health care, insurance, public attitudes and practices, etc. Think about how these affect your community.

3. Who are the stakeholders in these issues? Stakeholders are those individuals or organizations who have a “stake” in the issue. Some examples may include: Taxpayers,
community activists, community-based organizations, businesses, health departments, justice departments, survivors, researchers, etc. What makes them stakeholders and what do they have to gain from supporting or opposing health policies?

4. What are the policies that would impact these issues? For example, a policy could be a code of ethics by advertisers to eliminate tobacco advertising to children, stronger penalties for selling tobacco to minors, supporting mammography detection programs, etc.

5. Who has the power to advance the policies you have identified? Consider in particular policymakers at the local level in your community such as health departments, service providers, financial institutions, law enforcement, elected officials, insurance companies, professional organizations, trade associations, etc.

6. Barriers/Facilitators: What are the challenges to enacting your policy? Who are your allies? Who are your opposition?

Now that you have “mapped out” your issue, you can now begin to determine the data you need to:
- State the problem accurately in a compelling way;
- Show how your solution would alleviate the problem;
- Illustrate the negative consequences of not doing it;
- Refute arguments the opposition will use against you.

What message will the data deliver?
- Cost?
- Quality?
- Access?
- Equity?
- Rights?

What is the message that the data will deliver? What will mobilize your stakeholders and motivate the decision makers into action? Different things will convince different people. For some, the statistics alone with motivate them; for others a more personal story of overcoming hardships will be the key.

In either case, the statistics and the stories are most effective when they appeal to someone’s values. If you can supply data that not only accurately describes what people experience and data that appeals to their values or belief system, then you have some very powerful tools with which to achieve your goals.

Some common values that data can address are cost, quality, access, equity, and rights.
- Cost—what is the cost of the problem to taxpayers, community, business, individuals, and others.
- Quality—how is quality of life, environment, services, and programs impacted?
- Access—who has access to services, programs, insurance, jobs, education, clean air, etc.? Who doesn’t?
Building community capacity to collect data on health disparities

- **Equity**—is there an equitable distribution of resources among segments of a community?
- **Rights**—what are the rights of members of a community? What laws, regulations, or constitutional protections confer rights and on whom are the rights conferred?

What data does your audience need?

- Elected officials, juries, media, general public need data to understand the scope of the problem (the forest)
- Committee staffs, judges, special interest groups with legislative analysts need more specific information on who and what is impacted (individual trees)
- Agencies, courts, academics need details and statistics (roots)

The data you need not only depends on what you are trying to do, but also whom you have to convince. Some of this depends on the amount of time you have to present your position or the depth of information that you decide is appropriate. In general, the level of complexity you use when presenting data depends on the people you are trying to convince and their data needs.

**The Forest -- Big Picture:** Politicians, the general public and the media are audiences who tend to need information that is descriptive and easy to understand, often from an overall perspective or big picture point of view. The following is an example of “forest” type data from the Intercultural Cancer Council: American Indians and Alaska Natives have the poorest survival from “all cancers combined” as compared to any other racial/ethnic group (see http://iccnetwork.org)

**Individual Trees -- Some Details:** Committee staff, judges, and special interest groups with legislative analysts tend to need more detail than the big picture. These individuals want to know what kind of trees are in the forest or how many trees per square acre. They may want to know if they clear this part of the forest, what does that do to the ecosystem, etc. This information will have more layers to it; often the audience understands the general ideas, but does not understand the details. An example of data that provides some details about an issue is this one: The five-year relative survival rate for American Indian women with breast cancer continues to be the poorest of any racial group in the U.S. (Burhansstipanov et al. 1999; Burhansstipanov and Hollow 2001; Kaur p.c. 2000).

**Roots -- Specific Details:** Government agencies, court officials, and academic institutions often need data to be more academically focused or statistically driven to understand and critique. This type of data may require a high degree of accuracy because funding or planning decisions will be made based on the numbers. These audiences need to have as much detail as possible. Here’s an example of “roots” data: The five-year survival rate for breast cancer among American Indians throughout the U.S. appears to be between 40% (Kaur p.c. 2000) and 48% (National Cancer Institute 1992). This situation may be due in part to the fact that Native women are often diagnosed with advanced staged tumors (Li et al. Differences

Determining the Data You Need Brainstorming Activity:
Brainstorm about the data needed to address a community health issue. Create a list of community health issues of concern to your community. Identify one issue that is a priority. Next brainstorm about problems related to that issue that impact the community. Who has an interest in the issue and what data will they need to persuade them to be involved or counter their opposition? What policy solution is needed to address the issue and what data do you need to support your policy or program solution? Who are the policy makers who have the authority to adopt the policy and what data will they need?

Here's an example using the issue of childhood obesity. A blank worksheet to use in your brainstorm activity follows.

<table>
<thead>
<tr>
<th>Community Health Issue</th>
<th>Problem(s) associated with issue</th>
<th>Stakeholders (pro/con) in the Issue and data needed to persuade support, counter opposition</th>
<th>Policy/program Solutions And data that support solution</th>
<th>Policymakers with authority to adopt policy and data needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Childhood obesity</td>
<td>• Lack of food choices</td>
<td>• Schools: data on impact of obesity on academic performance, attendance</td>
<td>• Enable better food choices: cost data on missed attendance v. revenue gain from non-healthy foods</td>
<td>• Legislature, city councils, policy agencies, and school boards: local data on obesity impact, constituency (stakeholder) support for policy solution</td>
</tr>
<tr>
<td></td>
<td>• Inability to exercise during school day</td>
<td>• Food service providers: data supporting cost effectiveness</td>
<td>• Establish fitness programs: data showing association between fitness and academic performance</td>
<td>• Corporate voluntary policy: consumer support for healthy product choices</td>
</tr>
<tr>
<td></td>
<td>• Lack of safe play environments afterschool</td>
<td>• Commercial establishments: data demonstrating consumer demand for healthy food choices; concern about obesity influencing food choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Heightened risk of diabetes, heart disease, cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Determining the Data You Need Brainstorming Activity:
Brainstorm about the data needed to address a community health issue. Create a list of community health issues of concern to your community. Identify one issue that is a priority. Next brainstorm about problems related to that issue that impact the community. Who has an interest in the issue and what data will they need to persuade them to be involved or counter their opposition? What policy solution is needed to address the issue and what data do you need to support your policy or program solution? Who are the policy makers who have the authority to adopt the policy and what data will they need?

<table>
<thead>
<tr>
<th>Community Health Issue</th>
<th>Problem(s) associated with issue</th>
<th>Stakeholders (pro/con) in the Issue and data needed to persuade support, counter opposition</th>
<th>Policy/program Solutions And data that support solution</th>
<th>Policymakers with authority to adopt policy and data needs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from “Introduction to Health Data” and “Performing a Community Assessment”
Health DATA Program – Data, Advocacy and Technical Assistance
UCLA Center for Health Policy Research
You can begin your search for specific data by “cutting the data question.” That is, what specific data are needed? “Cutting the data question” provides you with basic questions to focus your search for data.

### Cutting the Data Question

When looking for data, consider the following questions before you begin your search:

- What is the **Problem/Issue** you are trying to resolve?
- What is the **Cause** of the Problem?
- What are the **Effects** of the Problem?
- What are the characteristics of the **Population**?
- Does **Geography** have an effect on the problem?

### Articulate the Primary Question to Be Answered

A **primary question** is a question you wish to answer with the information you collect to determine health disparities in your community.

<table>
<thead>
<tr>
<th>Primary Questions</th>
<th>Types of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Where do community residents go for health services?</td>
<td>Behaviors</td>
</tr>
<tr>
<td>2. What local health services do residents use the most? Which do they use the least?</td>
<td>Behaviors; Opinions</td>
</tr>
<tr>
<td>3. Are those services accessible to most people? (Are the location and hours of operation convenient? Is there public transportation to the site?)</td>
<td>Behaviors; Opinions; Services and resources provided; Policies</td>
</tr>
<tr>
<td>4. Which diseases or conditions affect the community the most?</td>
<td>Numbers or rates of disease, illness, disability, injury</td>
</tr>
<tr>
<td>5. Which of these diseases and conditions are best addressed by local health services and programs? Which are least addressed?</td>
<td>Services and resources provided; Utilization of services or resources provided; Opinions</td>
</tr>
<tr>
<td>6. How does our community compare to other communities on health issues?</td>
<td>Numbers or rates of disease, illness, disability, injury</td>
</tr>
</tbody>
</table>
7. What kinds of associations, networks, clubs and other volunteer groups exist in the community? Level of awareness and knowledge

8. What kinds of community issues have these associations taken on in the past? Level of awareness and knowledge

9. What current community issues are the most important for residents? With which would residents most likely get involved? Level of awareness and knowledge; Opinions and priorities

10. Where do teens in the community get information regarding diet and nutrition? Behaviors

11. What do high school students know about eating a healthy diet? Level of awareness, knowledge and beliefs

12. What foods are available for high school students, both on campus and nearby? Maps; Photographs; Sales transactions

## Sources of Data - Secondary Data

Secondary data is data collected by other agencies and research institutions. Examples of secondary data include the following. Check with similar sources in your local community:

### 1. State agencies and foundations:
- California Cancer Registry: [http://www.ccrcal.org/](http://www.ccrcal.org/)
- Office of Statewide Health Planning and Development, Health Care Information Division: [http://www.oshpd.ca.gov/hid/index.htm](http://www.oshpd.ca.gov/hid/index.htm)
- RAND Corporation Public Use Databases: [http://www.rand.org/services/databases.html](http://www.rand.org/services/databases.html)

### 2. State surveys:
- California Cancer Registry: [http://www.ccrcal.org/](http://www.ccrcal.org/)
- California Health Interview Survey, Ask CHIS site: [http://ww.chis.ucla.edu/main/default.asp](http://ww.chis.ucla.edu/main/default.asp)

### 3. Federal agencies:
- Centers for Disease Control and Prevention, CDC Wonder: [http://wonder.cdc.gov/](http://wonder.cdc.gov/)
- Centers for Disease Control and Prevention, AIDS Public Use Data by major metropolitan area: [http://wonder.cdc.gov/AIDSPublic.html](http://wonder.cdc.gov/AIDSPublic.html)
4. **Federal surveys through the National Center for Health Statistics:**
   http://www.cdc.gov/nchs/
   - National Health Interview Survey (NHIS):  
     http://www.cdc.gov/nchs/products/elec_prods/subject/nhis.htm
   - National Health and Nutrition Examination Survey (NHANES):  
     http://www.cdc.gov/nchs/nhanes.htm
   - Behavioral Risk Factor Surveillance System (BRFSS):  
     http://www.cdc.gov/brfss/
   - Youth Risk Behavior Surveillance System (YRBSS):  
     http://www.cdc.gov/nccdphp/dash/yrbss/index.htm
   - National Immunization Study public use data files:  
     http://www.cdc.gov/nis/datafiles.htm
   - Ambulatory Health Care Data:  
     http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm#Micro-data
   - National Hospital Discharge and Ambulatory Surgery data:  
     http://www.cdc.gov/nchs/about/major/hdascd/nhds.htm
   - National Nursing Home Survey data:  
     http://www.cdc.gov/nchs/about/major/nnhsd/nnhsd.htm
   - National Home and Hospice Care Survey data:  
     http://www.cdc.gov/nchs/about/major/nhhcsd/nhhcsd.htm
   - National Employer Health Insurance Survey data:  
     http://www.cdc.gov/nchs/about/major/nehis/nehis.htm
   - National Health Provider Inventory data:  
   - National Survey of Family Growth data:  
     http://www.cdc.gov/nchs/nsfg.htm
   - State and Local Area Integrated Telephone Survey:  
     http://www.cdc.gov/nchs/slaits.htm
   - National Vital Statistics System:  
     http://www.cdc.gov/nchs/nvss.htm
2. Finding Data

It is important to be as specific as possible about the issue or problem before starting your data search. Your health focus, population of interest and geographic area will help you conceptualize the different types of data that you will need and where you may go to access the data.

Data search challenges

There are three themes that may arise during your data search.

1. **Limited Resources:** Organizations may lack the internal capacity to search for appropriate data or fully utilize the data that they find. Often the organizations may not have the time, knowledge base, and/or people power to access the data.

   Plan to spend time looking for data when you do a campaign. Be realistic about how much time this is going to take you and the manpower and expertise that you will need. Getting the data you need can take weeks, or months. You should plan accordingly.

2. **Access:** The limited quantity or non-existence of certain data can be very problematic. It may take some time to find the data you need or it may require that you create a local estimate based on national or state data you find. See section below on how to make estimates using national or state data.

3. **Quality:** It can be difficult to determine the reliability of information, especially if time constraints are involved in the process. Refer to the five criteria for evaluating data (credibility, specificity, generalizability, reliability and timeliness).

What can you do if the data you need are not available?

Sometimes the data you need may be limited or unavailable for your specific population. In these cases, you can try to localize data, or take existing data and show how it applies to your population of interest. The following approaches can help you localize data:

A. Use proxy measures
B. Extrapolate national or state data to make local area estimates
C. Paint a picture
D. Ask a researcher

A. Using Proxy Measures:

"Proxy" measures are information that can substitute for the data you need because it is closely related to your issue. For example, you may need recent elder poverty data for your neighborhood. Since Medi-Cal is limited to low-income persons, you could take the number of elders receiving Medi-Cal as an indicator or proxy of the poverty level of elders in the neighborhood (this would be an undercount since not all low-income elders are on Medi-Cal).
Similarly, you could use the number of emergency room visits for falls among older persons as a proxy of risk for all falls by the elderly. These data do not give you a precise number or rate of your problem, but the data provide useful comparisons between communities, such as, “The elderly rate of poverty in our neighborhood may be much higher than the city average, as shown by our higher rate of elders on Medi-Cal.”

A major advantage of this approach is its low cost. The data can be relatively easy and inexpensive to collect. However, there are some concerns with bias. Your estimates may be biased because they are not able to capture actual rates or precise numbers but it does offer some estimates for when data are not available.

B. Extrapolating From Existing Data:
Extrapolation involves taking a national or state pattern of a problem and applying that pattern to your local area. Diabetes, for example, is a condition that needs on-going medical care. There are large differences by race and age in the rates of diabetes. To estimate the number of persons with diabetes in your community based on national trends (you could also take state or county data if you have access to it), you can take the following steps:

1. Identify the diabetes rate (percent with diabetes) using the national data source. Obtain the rate for subgroups where there is variation (e.g. race, sex, age, income) For example, the diabetes rate for Latinos nationally is .02 for age 18-44, .143 for age 45-64, and .204 for age 65 and over. [Source: National Health Interview Survey, Table 8 at http://www.cdc.gov/nchs/fastats/pdf/sr10_209.pdf]

2. Identify the number in the population for the same subgroups locally. For example, say your community has the following:
   a. 30,000 Latinos ages 18-44,
   b. 11,000 Latinos ages 45-64, and
   c. 2,000 Latinos ages 65 and over.

3. Multiply the national rates by the local numbers and add them up.

\[
\text{Number of Latinos with diabetes in your community} = \sum (\text{National Rates} \times \text{Population in your community (by the different ages)})
\]

\[
0.02 \times 30,000 \text{ (ages 18-44)} = 600 \\
0.143 \times 11,000 \text{ (ages 45-65)} = 1,573 \\
0.204 \times 2,000 \text{ (ages 65 and over)} = 406
\]

Next add up the various populations with diabetes:

\[
600 + 1,573 + 406 = 2,579
\]

There are an estimated 2,500 Latinos with diabetes in your community.
This method does not provide “precise” data, but it offers a way to generate useful estimates that can be used in your program planning and policy advocacy work to address disparities.

C. Paint a Picture

If you are unable to find the exact numbers you need to describe the impact of a health issue in your community. You can PAINT A PICTURE with the information you do have. The approach allows you to piece together data from several sources to illustrate your argument. It also allows you to use anecdotal information that can compliment your statistical data.

D. Ask a Researcher

- If you find a particularly helpful study, it might be possible to contact the researcher to find out more.
- Expect that it will take time, many researchers have moved on to their next discovery by the time data from their last experiment becomes available to the public.
- When you do get data this way, pay attention to any caveats the researcher places on the data, these caveats may be the reason the researcher did not publish that information, even if he/she found it interesting.
- Seek out those sources of information that provide ongoing support or technical assistance.

Important Notes:

- Obtain data from a CREDIBLE source that resembles/approximates the data you need.
- Consider the TIMELINESS and GENERALIZABILITY of the data. Note the similarities and differences between the demographics of the data and the demographics of your constituents.
- What will the estimate be used for? Is it appropriate to estimate in certain instances?
- Sometimes no number is better than a bad one; sometimes a fuzzy one is better than none. You must decide.
- Be prepared to defend the information you use and your methods for getting it.

Combining Quantitative and Qualitative Data

Keep in mind that there are two different kinds of data you may encounter through your data search, quantitative and qualitative.

Quantitative data are usually measured and expressed in the form of numbers or percentages. This data answers who, what, when and where.

Qualitative data are usually measured and expressed in the form of words, concepts, themes, or categories rather than numbers. Qualitative data is often used to gain a more in-depth understanding of a particular incident or phenomenon—answering how or why something is occurring.
3. Collecting Your Own Data

Now that you have identified what data is needed to answer your questions and have identified potential secondary data sources, you may determine what data, if any, you need to collect from scratch. This is called primary data. The most difficult part about selecting the most appropriate data collection method is making sure that your partnership has the resources to perform that particular method. The following chart discusses the pros and cons of each method.

### Data Collection Methods Chart

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asset Mapping</td>
<td>• Builds on existing community assets&lt;br&gt;• Can generate a lot of community participation&lt;br&gt;• Mapping the inventory creates a visual depiction of existing and lacking assets&lt;br&gt;• Data can be used to raise awareness about the availability of assets, develop or improve services and programs, or to apply for funding</td>
<td>• Finding the right maps can be difficult, and mapping software can be expensive and difficult to use&lt;br&gt;• Some community assets will be difficult to map if they don’t have a physical location&lt;br&gt;• Needs community buy-in and collaboration to adequately inventory up-to-date community resources</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>• Flexible&lt;br&gt;• Captures rich, in-depth data&lt;br&gt;• Immediate results&lt;br&gt;• Encourages and stimulates individuals to share more openly&lt;br&gt;• Data can be combined with quantitative data to provide a complete picture about an issue</td>
<td>• May be challenging to recruit participants&lt;br&gt;• Need to schedule at least 2-3 focus groups to capture diversity&lt;br&gt;• Difficult to generalize results to the larger population because of small numbers of participants&lt;br&gt;• Difficult to compare results across groups</td>
</tr>
<tr>
<td>Survey</td>
<td>• Data can be collected from a lot of respondents easier than any other method&lt;br&gt;• Can get a large enough sample that can be representative of the larger population&lt;br&gt;• Findings can be generalized to the larger population&lt;br&gt;• Can cover a lot of topics&lt;br&gt;• Can easily compare different groups’ data to each other</td>
<td>• Survey instrument must be carefully constructed to avoid leading questions, and to make sure the appropriate responses are available&lt;br&gt;• Response rates can be low for self-administered surveys, especially mailed ones&lt;br&gt;• Response will be low if survey is too long</td>
</tr>
</tbody>
</table>
## Data Collection Methods Chart

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| **Key Informant Survey**                    | • Detailed and rich data can be gathered in a relatively easy and inexpensive way  
• Allows interviewer to establish rapport with the respondent  
• Provides an opportunity to build or strengthen relationships with important community informants and stakeholders  
• Can raise awareness, interest, and enthusiasm around an issue  
• Can contact informants to clarify issues as needed | • Selecting the “right” key informants may be difficult so they represent diverse backgrounds and viewpoints  
• May be challenging to reach and schedule interviews with busy and/or hard-to-reach respondents  
• Difficult to generalize results to the larger population unless interviewing many key informants |
| **Community Forums**                        | • Can raise awareness and knowledge on an issue  
• Relatively easy and inexpensive to conduct  
• Allows for community and stakeholder participation in issue  
• Participants can identify new areas or topics not previously considered  
• Forums can help identify new leaders or stakeholders that may help in the planning and implementation of a project or initiative | • May be difficult to schedule a series of public meetings with the community members and other stakeholders you want to reach  
• Participants may not be representative of the larger population as those who attend may not reflect the entire community or target audiences  
• Participants come with their own expectations and agendas |
| **Direct Observation and Photography (Rapid Appraisal Methods)** | • Relatively easy and inexpensive  
• Data can be gathered quickly  
• Can create community input and participation  
• Provide descriptions and visual imagery that give meaning to quantitative data | • Can be difficult to interpret and summarize photographs and observation notes  
• Can be difficult to represent the entire community experience |
# Planning Worksheet: Questions, Data Types and Data Sources

<table>
<thead>
<tr>
<th>Primary Questions</th>
<th>Types of Data</th>
<th>Source of Data</th>
</tr>
</thead>
</table>
| Example: Who has cervical cancer in our community? (What is their age, race/ethnicity, insurance status, income level, language preference, number of years living in US, etc.?) | • Cervical cancer diagnoses  
• Demographics  
• Behaviors (regular screening practices, sees a doctor regularly, etc.) | • Community women  
• Local clinics  
• Family planning outreach and education programs  
• Department of public health          |

1.

2.

3.

4.

5.

6.

7.

8.