

Introduction to Health Data Using Online Resources: AskCHIS Training Workshop

Training Manual



UCLA Center for Health Policy Research
Health DATA Program: Data. Advocacy. Training. Assistance.

SUGGESTED CITATION:

LESSARD, L, PARKS, A., TOY P. *Introduction to Health Data Using Online Resources: AskCHIS Training Workshop*, LOS ANGELES: UCLA CENTER FOR HEALTH POLICY RESEARCH, HEALTH DATA PROGRAM, JUNE 2013.

THE VIEWS EXPRESSED IN THIS CURRICULUM ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REPRESENT THE UCLA CENTER FOR HEALTH POLICY RESEARCH, THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, OR KAISER PERMANENTE.



THE UCLA CENTER FOR HEALTH POLICY RESEARCH IS BASED IN THE UCLA FIELDING SCHOOL OF PUBLIC HEALTH AND IS AFFILIATED WITH THE UCLA LUSKIN SCHOOL OF PUBLIC AFFAIRS.

VISIT THE CENTER'S WEBSITE AT: WWW.HEALTHPOLICY.UCLA.EDU

THE HEALTH DATA PROGRAM IS A PUBLIC SERVICE OF THE UCLA CENTER FOR HEALTH POLICY RESEARCH.

VISIT THE HEALTH DATA WEBSITE AT: WWW.HEALTHDATA.UCLA.EDU



DEVELOPMENT OF THE ASKCHIS COMMUNITY WORKSHOP WAS MADE POSSIBLE THROUGH A GRANT FROM THE CALIFORNIA ENDOWMENT AND KAISER PERMANENTE



ACKNOWLEDGMENTS:

The authors would like to acknowledge contributions from Health DATA Principal Investigators Steven P. Wallace, PhD and Roberta Wyn, PhD, and the contributions of Victoria Ojeda, PhD and Jennifer Erausquin, MPH who developed the initial AskCHIS Community Workshop curriculum. We also acknowledge support received from Zero Divide/Community Technology Foundation of California and The California Endowment to develop the first edition of the *Online Data Access Training for the California Health Interview Survey: The AskCHIS Community Workshop Workbook*

***UCLA Center for Health Policy Research
Health DATA Program***

**Introduction to Health Data Using Online Resources:
AskCHIS Training Workshop**

Table of Contents	Page
A. INTRODUCTION: AskCHIS In Person Training Overview	1
Project Summary	
Learning Goal and Objectives	
Overview of Today's Training	
Technical Assistance	
Background of the Health DATA Program	
B. LEARNING MODULE 1: Data for Decision Making & Developing and Answering Data Questions	7
Part I: Understanding Data	9
Learning Objectives	
Introduction to Data	
Evaluating Data: Determining if a Data Source Will Meet Your Needs	
Strengths and Limitations of CHIS as a Data Source	
Part II: Developing & Answering Data Questions	15
Learning Objectives	
Addressing Community Health: Formulate and Answer Data Questions	
CHIS Case Studies	
C. LEARNING MODULE 2: Accessing Data from the AskCHIS Online Data Query System	21
Part 1. Introduction to AskCHIS	22
Learning Objectives	
What is a Query?	
CHIS Data Highlights	
Using AskCHIS	
Part 2. Developing the Data Query	32
Learning Objectives	
What can AskCHIS Provide?	
AskCHIS Data Resources	
Data Resources for AskCHIS Users	
Defining the Data Query & Population and Preparing the Data Query	
Part 3. Executing Queries & Retrieving Data	36
Executing Sample Queries	
Getting Results	
Part 4. Interpreting and Presenting AskCHIS Data	47
Appropriate Analysis of CHIS Data	
Identifying Solutions for Stabilizing Estimates in Small Populations	

Table of Contents (Continued)

Appendix A: Health DATA Toolkit	52
Combining Quantitative and Qualitative Data	
Glossary of Useful Data Terms	
In Focus: Confidence Interval	
What to Do When the Data You Need Is Not Available	
Appendix B: AskCHIS Worksheets and Examples	65
<i>Worksheet 1: Defining the Data Query & Population and Preparing the Data Query</i>	
<i>Worksheet 2: Query for a Two-Variable (Bivariate) Table</i>	
<i>Worksheet 3: Interpreting and Applying the Data Results</i>	
<i>Worksheet 4 Strategies for Handling Unstable Estimates</i>	
Appendix C: Data Resources	75
List of Data resources- in alphabetical order by topic – which you can refer to for additional information to help you search for, understand, and apply relevant data.	

A. AskCHIS Training Workshop Overview

A. AskCHIS Training Overview

Project Summary:

The **Introduction to Health Data Using Online Resources: AskCHIS Training Workshops** are a project of the Health DATA Program of the UCLA Center for Health Policy Research. These trainings are funded by a grant from The California Endowment.

There are 2 types of trainings available, in person (approximately 6-hours in length) and online (approximately 2 hours in length.) Both trainings discuss available data sets, illustrates the use of AskCHIS, provides activities and worksheets, and hands on use of the Internet and computer software to familiarize participants with data and build critical thinking skills to address their data needs. Once their data needs are identified, participants work to obtain and apply relevant data. Technical assistance is provided by project staff following the training to assist workshop participants with their use of tools learned in the training.

Learning Goal and Objectives:

Goal:

The goals of the AskCHIS Trainings are to train and provide technical assistance to staff of Community Based Organizations and local health departments who are planning to use California Health Interview Survey (CHIS) data to increase their capacity to obtain and apply health and social data using computer-based and Internet-based tools.

Objectives:

Upon completion of this training, you will be able to:

- To determine the appropriate uses and limitations of CHIS data
- Identify relevant and appropriate data sources to answer data questions;
- Increase knowledge and skills to use the AskCHIS data reporting system and computer tools to access, interpret, and apply data from the California Health Interview Survey

Overview of Today's Training:

Today's workshop will provide training on how to understand data, identify and evaluate data sources and access and analyze data from the AskCHIS online data query system.

The workshop curriculum contains two modules:

- Understanding Data and Developing and Answering Data Questions;
- Accessing Data from the AskCHIS Online Data Query System

1. Data For Decision Making

This training module consists of two sessions to help you build your knowledge of data and help you develop questions that can be answered using data.

Understanding Data

This session provides an overview of the benefits and limitations of data, different data types and sources and criteria for evaluating data sources.

Developing and Answering Data Questions

This session helps you develop, answer and present on data questions using appropriate data sources.

2. Accessing Data from the AskCHIS Online Data Query System

This is a five part hands-on, step-by-step training module for accessing data from the California Health Interview Survey through its online data query system, AskCHIS. The below five sections will help you learn how to construct a query, or a search statement, to get data from CHIS. You will also learn how to export and analyze data as well as creating a brief synopsis of your findings.

Part I. Introduction to AskCHIS

Part II. Developing the Data Query

Part III. Executing Queries & Retrieving Data

Part IV: Additional Features

Part V. Interpreting and Presenting AskCHIS Data

Technical Assistance:

You can receive technical assistance from the Health DATA Program following today's workshop and request one-on-one help with the information presented in today's workshop.

To request technical assistance, contact the Health DATA staff at (310) 794-0950 or email us at hdp@ucla.edu

Background of the Health DATA Program:

The **UCLA Center for Health Policy Research** conducts research on national, state, and local health policy issues. The research, service, and education programs of the Center emphasize a community- and population-based perspective to improve health outcomes.

The Center provides training and technical assistance through the **Health DATA, Data, Advocacy, and Technical Assistance Program**. The Health DATA (Data, Advocacy, and Technical Assistance) program was created in 1997 in response to community needs for data and the skills to use the data effectively. The Health DATA program builds the knowledge and skills of organizations to address community needs by enhancing their capacity to effectively apply data in their decision-making, planning, policy advocacy, and other activities.

The **California Health Interview Survey (CHIS)**, based at the Center, is the largest state health survey ever conducted in the United States, collecting information from over 40,000 households drawn from every county in the state. CHIS provides statewide estimates for California's overall population, including important information not previously available on an array of racial/ethnic populations, and local-level estimates for counties with populations of 100,000 or more. CHIS data is used for local planning and for making comparisons across counties.

B. Data for Decision Making & Developing and Answering Data Questions

Using *AskCHIS* to Promote Community Health

B. Data for Decision Making & Developing and Answering Data Questions

Part 1: Understanding Data

I. Learning Objectives

- Distinguish between different data types;
- Identify the type of data CHIS provides;
- Review how to evaluate data sources;
- Review benefits and limitations of CHIS data;

II. Introduction to Data

What is data? Data is factual information used for decision-making.

Data can take many forms and can be categorized into four major categories: *quantitative/qualitative* data and *primary/secondary* data.

- *Quantitative* data is numerical (e.g. the number of health care facilities in a city)
- *Qualitative* data is textual/words (e.g. an individual's observations and experiences of a particular health care facility).
- *Primary* data is collected and analyzed directly by individuals, communities and researchers (e.g. a survey developed, administered and analyzed by hospital staff that rates the quality of patient care)
- *Secondary* data is collected and provided by a third party such as a non-profit, university or government agency (e.g. US Census).

The matrix below provides examples of data within each of the four data categories (i.e. quantitative/qualitative and primary/secondary). The matrix draws a distinction between *how* data is collected (i.e. primary vs. secondary data) and *what* data describes (i.e. qualitative vs. quantitative). The matrix is entitled *Data Landscape* because every type of data fits within and can be described by the four categories presented.

DATA LANDSCAPE			
<i>What data describes</i>			
<i>How data are collected</i>		Qualitative <i>Textual/words data</i>	Quantitative <i>Numerical data</i>
	Primary <i>Data collected and utilized directly by individuals, communities and researchers</i>	<ul style="list-style-type: none"> • Data collected by hospital staff via a focus group on patient's opinions of care quality • Case study conducted by community members on resident's experiences with financing health insurance 	<ul style="list-style-type: none"> • Hospital staff count and analyze the number of HIV positive patients admitted • Community members record the number of traffic fatalities on major streets.
	Secondary <i>Data collected and provided by a third party such as a non-profit, university or government agency</i>	<ul style="list-style-type: none"> • Report published by a non-profit organization that provides testimonials on immigrants' experiences accessing health care in the United States • Case studies produced by a university that highlight emergency room best practices 	<ul style="list-style-type: none"> • AskCHIS • US Census data • Scorecards produced by non-profit organizations

AskCHIS is secondary data because it is collected and produced by a third party, the UCLA Center for Health Policy Research. Also, AskCHIS is a quantitative database that provides numerical statistics on a variety of public health topics statewide and by county. Today's workshop will focus on obtaining and analyzing secondary, quantitative data from AskCHIS.

For additional information on different types of data, how they are collected, and how they can be used, please see Appendix A: The Health DATA Toolkit.

III. Evaluating Data Sources: Determining if a data source meets your data needs

It is important to evaluate any secondary data source before you decide to use it. You can do this by thinking about secondary data in terms of the following five evaluation criteria. Since we will be discussing CHIS today, the examples below apply these evaluation criteria to CHIS.

1. Credibility - Who produced the data?

Credibility refers to the source of the data. When considering a data source, you may ask yourself: What is the reputation of the data source? Can you trust the research entity that produced the data? What is the entity's mission, public image or reputation? Do organizational interests bias the interpretation and presentation of the data?

It is important to note that data sources have different levels of credibility based upon the interest to data collector in presenting the data in a certain light. For example, official government statistics are highly credible in comparison to industry group data that always supports their financial interests (e.g. tobacco institute). University research centers are similar to government statistics, in part because they undergo the scrutiny of other researchers.

Example: CHIS data maintain a high level of credibility as it is collected and produced by the UCLA Center for Health Policy Research, a research unit within the UCLA School of Public Health. The Center is one of the nation's leading health policy research centers and the premier source of health policy information for California. The Center's mission is to improve the public's health by advancing health policy research, public service, community partnership and education. CHIS is funded by a variety of philanthropic and governmental organizations that promote the well-being of Californians and knowledge of health issues.

2. Specificity - Does the data capture what I want to measure?

Specificity refers to how well the data capture what you want to measure. You may want to ask yourself: How do the goals or content of the study relate to your needs? Are the conditions of the study unique to a particular case? How close is the relationship between what you need and the research data?

Example: CHIS data captures a wide variety of health topics and demographics throughout California. When searching CHIS, you may ask yourself, "Does the data really capture what I want to know?" Be sure to examine each CHIS variable closely to ensure its specificity to the issue you want to address and note that it may not provide the data you want. Below is an overview of the type of information CHIS provides:

- **Respondent Characteristics:** Age, gender, race and ethnicity, marital status, sexual orientation, language spoken, citizenship/immigration status, country of birth;
- **Geography:** CHIS data are collected from 44 geographic areas across California that represent 41 individual counties and three groupings of counties with smaller populations;
- **CHIS 2007 Topics:** health status, health conditions, mental health, health behaviors, women's health, cancer history and prevention, dental health, food insecurity/hunger, food environment, neighborhood and housing, access to use of health care, health insurance, public program

eligibility, interpersonal violence, parental involvement, child care, employment, and income.

3. Generalizability - Is the data applicable from one population to another population?

Generalizability refers to specific data and how well it can be used to describe other populations. You may want to ask yourself: What are the characteristics of the secondary data's population/sample? Do the participants provide data that can be applied to other similar populations or sub-populations? Do the "who, what, why, when and where" of the data relate to the "who, what, why, when and where" of your work? For example, how well can information collected across California accurately describe the residents of your county?

Example: CHIS data are very generalizable at the state and county levels due to rigorous sampling techniques that ensure a sufficient amount of data to describe populations throughout the state. For instance, CHIS data are collected every two years via telephone interviews with upwards of 50,000 California households participating. CHIS sets minimum target numbers for each geographic area to ensure a statistically representative sample of the state's diverse population.

Additionally, CHIS uses many techniques to interview enough people from several ethnic groups to better characterize most major and minor racial and ethnic populations statewide. During each survey cycle, thousands of CHIS interviews are conducted in languages other than English and certain ethnic minority groups, such as American Indian and Asian subgroups, have been oversampled in the past to ensure a representative sample.

4. Reliability - How was the data collected?

Reliability refers to the accuracy of the data. You may want to ask yourself: Can the data be trusted to be accurate? Does the research seem free of bias or error? Have the methods and results been proven? Was the research repeated? If so, did the second study get the same or similar results?

Example: CHIS data are very reliable as the highest research practices and standards are applied. A few of the methods used to ensure data reliability are mentioned below:

- A large number of individuals are surveyed to ensure a representative sample of the entire state's diverse population;

- To attain a fair and unbiased sample, computers randomly draw telephone numbers from each geographic area sampled with a minimum number of people to include;
- To ensure that public health trends are accurately recorded over time, CHIS selects and surveys unique households during each survey cycle. CHIS 2007 data also includes a sample of cell-phone numbers;
- CHIS randomly selects and interviews only one adult per household. Only that selected person can participate in each household;
- If there are minor children in the household, CHIS also asks questions about the adolescents and younger children;
- To ensure the inclusion of California's diverse racial and ethnic populations, thousands of CHIS interviews are conducted in languages other than English.

5. Timeliness - When was it collected?

Timeliness refers to when the research was conducted. You may want to ask yourself: When or how recently was the data collected? Have major changes occurred in your population of interest, geographic area, or disease topic that may not have been captured by this data source? Is it the most recently available data that will suit your needs?

Example: CHIS data cycles occur frequently and are conducted every two years since 2001. 2007 CHIS data will be made available in 2008. The high frequency of collection and production of CHIS data increases the relevancy and timeliness of data findings and trends.

IV. Strengths and Limitations of CHIS as a Data Source

If you are unsure about how the criteria above apply to a secondary data source, then go to the source and ask them, if possible.

No data are perfect. No data will fulfill all of the above criteria. It is up to you to balance the pros and cons of each data source and decide what criteria are more and less important for your needs.

Example: Based on the evaluation criteria reviewed above, CHIS data are very useful for making powerful statements about community health. However, CHIS data do have limitations. Below are examples of questions CHIS can and can *NOT* answer:

- Example questions that CHIS data can answer:
 - What is the estimated number of people in my county with health insurance? (CHIS collects data by county);

- What is the overall health status of my county in comparison to the state overall? (CHIS can compare information between counties and the state);
 - Are there differences in the number of people that smoke in my county among different ethnic groups and has it changed over time? (CHIS can compare information across ethnic groups and over time.
- Example questions that CHIS can *NOT* answer:
 - How has the rate of teens having sex changed over the past 10 years statewide? (CHIS only collects data after 2001 and data is currently available through 2009);
 - How does the number of women with breast cancer compare with women from other states and nationally? (CHIS only collects data on California households).

B. Data for Decision Making

Part 2: Developing and Answering Data Questions

I. Learning Objectives

- Explain steps to develop and answer data questions
- Explain how CHIS can help answer data questions

II. Addressing Community Health: Formulate and Answer Data Questions

Now that you have a better understanding of what data are, the different types of data, the criteria to evaluate data sources and terms commonly used when working with data, let's explore how to use data to better understand and describe community health and advocate for change.

The four steps outlined below will help you formulate and answer data questions regarding community health. Each step includes an example using CHIS data to illustrate how data can be used to complete each task.

1) *Identify a question that can be answered using data*

Policy makers, advocates and community members often have important questions about their population of interest, its health and health resources. A clear definition of the data question's scope and the issues addressed can increase the likelihood that data sources and other resources are properly identified and applied to produce an answer.

Consider the population, location and pervasiveness of a specific community health issue when developing a data question. A careful review of these factors can ensure that the question's issue and its parameters are clear, easily communicated and that relevant data sources are gathered. Please note that an unclear data question can lead to a lack of clarity when determining what data could answer it. This often leads to potentially tedious and costly data collection and analysis, as well as the misallocation of scarce resources.

Example: CHIS data can help identify a data question. Suppose the ABC Advocacy Group wants to raise awareness around pediatric asthma in Southern California. Due to limited resources, ABC must first select a county to target their efforts before initiating advocacy work. ABC could use CHIS data to select a county whose incidence of pediatric asthma merits further attention. In this case, the data question is, "Which

Southern California County has the highest reported incidence of pediatric asthma?"

CHIS data gives insight into the prevalence of pediatric asthma and its determinants over time, geographic concentration and pervasiveness within sub-populations. These data inform the advocacy group about potential target populations, location, concentration and need. Based on the results produced, CHIS can also help determine if additional data is needed to make a more informed population selection. In this example, CHIS data informs the data question and assists the ABC Advocacy Group to locate their population of interest.

2) *Identify appropriate data types and sources*

As discussed in Section B, Part I of this workbook (Understanding Data), certain types of data are more useful in answering certain questions. For example, *quantitative* data is useful when questions require an answer in the form of a number, rate, or percentage, such as the number of inoculations provided per month or the rate of infection of a particular disease. Conversely, *qualitative* data is useful when questions require words or descriptions to answer such as patient's descriptions of their own opinions, feelings or experiences. words or descriptions to be answered such as patient's opinions, feelings and experiences. Also, *secondary* data is useful when time limits and other resource restrictions prevent *primary* data collection which can be costly and labor intensive. However, *secondary* data may not provide the level of credibility, generalizability, specificity or reliability that *primary* data provides to accurately answer the question.

Example: When identifying and selecting appropriate data sources to answer your data question, it may be helpful to work backwards by considering the type of answer you need before data sources are identified. Using the ABC Advocacy Group example, the organization's goal was not only to identify a target population but to justify and explain the type of advocacy work to conduct.

In order to answer their data question, ABC will need detailed information about existing programs, advocacy efforts and the cost of these efforts to determine how useful investing its resources will be. By first considering the answer's key components (e.g. The ABC Advocacy work will conduct a _____ advocacy campaign on pediatric asthma in _____ county with _____ population, that reaches _____ number of children over _____ time period), the types and sources of data needed become clearer.

To answer the question above, several data types and sources are needed including information on the population's size (quantitative), a justification of the type of advocacy work to be conducted (qualitative) and a time period (quantitative). CHIS data (secondary) could be used to provide information on the population and its size while discussions with local health providers (primary) can justify the type of advocacy work needed and the time period.

3) *Identify stakeholders that can help answer the data question*

After considering the type of data and sources needed to answer a question, multiple data sources are often required. Acquiring these data can involve connecting with local, community stakeholders that represent, work with or are a part of the community of interest. These individuals and organizations may provide or have access to unique resources that can help answer the data question. These resources can be categorized into material and non-material resources:

- Non-material: professional expertise working with and access to the target population, data and research experience, political and professional influence or allies, the ability to provide oversight and guide the inquiry, etc.;
- Material: funding, data, technology, facilities (e.g. hospital, schools, day care centers), health care professionals, clients/patients, etc.

Both material and non-material resources are often needed to access data, particularly when primary data must be collected from patients or community members. In these cases, a stakeholder who is a health provider may provide you with the opportunity to collect data from your population of interest (non-material), or provide you with access to patient trends (material).

Example: In order to find the data ABC Advocacy Group needs to answer their data question, ABC will need to identify and partner with local stakeholders such as the department of public health, hospital administrators, and local advocates. These stakeholders can help obtain the data needed to answer their data question, in order to justify and focus ABC's work within the community.

4) *Present the data*

Once the appropriate data have been identified and acquired from the appropriate data sources, the data should be analyzed and presented to answer the question. Below are some of the most common ways quantitative and qualitative data are analyzed and presented:

- quantitative data: tables, pie charts, bar graphs;
- qualitative data: case studies, testimonials, focus group summaries.

Presenting data in a simple and consistent way is important because it allows you to communicate your findings clearly and present the answer(s) to your data question. For an overview of how to analyze and present quantitative data, please see Presenting Data in Appendix D: Data Resources.

Example: AskCHIS has the capacity to produce rates, data tables, pie charts and bar graphs for all variables within the database. You can also download the data into an Excel spreadsheet which can produce more sophisticated analysis than AskCHIS. CHIS data can also be downloaded and analyzed using common statistical analysis software such as STATA and SAS.

Brainstorming Activity: Formulate and Answer Data Questions

Complete the chart below for a specific issue in your community. Consider how CHIS can help accomplish each step.

Community Health Issue	Data Question	Data Types Needed	Data Stakeholders	Data Presentation

III. CHIS Case Studies: Examples of Community Health Decision Making Using CHIS

The following case studies illustrate how statewide, regional and county organizations have used CHIS to describe community health and how data was used for decision-making.

California Health Care Reform:

Both California Governor Schwarzenegger's office and Democratic legislators used CHIS data to develop health care reform proposals. Governor Schwarzenegger also held a press conference at UCLA in which he cited a cornerstone CHIS publication - *The State of Health Insurance in California* - as important evidence of the need for health care reform. California State Senate Health Committee staff used CHIS

data to assess coverage effects of each proposal. The Legislative Analyst Office used CHIS data to analyze the Governor's health care reform proposal for state legislators and for the report, *The 2007-08 Budget: Perspectives and Issues*. Advocacy groups on all sides of the issue also relied on CHIS data to inform their positions and craft their arguments.

First 5 County Commissions:

More than a dozen First 5 county commissions - created by California voters to direct tobacco tax revenues to promote early childhood development - have used CHIS data to develop new public-private expansion programs for children ineligible for private insurance, Medi-Cal or Healthy Families. In most cases, CHIS has been the only data source available. First 5 agencies from Fresno, Kern, Los Angeles, Riverside, San Joaquin, Marin, San Luis Obispo and Tulare counties have relied on CHIS data to advocate for and plan these expansion projects, as has a coalition of commissions from Sacramento, Colusa, El Dorado, Yolo and Yuba counties.

Increased access to food stamps for poor families:

A policy brief on food insecurity by the UCLA Center for Health Policy Research and the California Food Policy Advocates prompted then-Assemblyman Darrell Steinberg to introduce AB 231 to increase participation in the federal food stamp program.

C. Accessing Data from the AskCHIS Online Data Query System

C. ACCESSING DATA FROM THE AskCHIS ONLINE DATA QUERY SYSTEM

This section of today's workshop focuses on accessing data from the California Health Interview Survey using an online tool called AskCHIS. There are five parts to this section:

1. Introduction to AskCHIS
2. Developing a Data Query
3. Executing Queries & Retrieving Data
4. Downloading & Manipulating Data in Excel
5. Interpreting and Presenting AskCHIS Data

Part 1: Introduction to AskCHIS

Learning Objectives:

- Understand what a query is.
- Understand how AskCHIS can help you access data about Californians.
- Become familiar with the basic steps in using AskCHIS.

What is a query?

Query: a search or inquiry.

Many people use the Internet to get information. When you use an Internet search engine such as Yahoo or Google to request information and then obtain some results, you are performing a **query**. You are searching for data that someone else has collected. This session will introduce you to one system for internet-based data queries, **AskCHIS**. You will also work on creating a query that will give you the data that you need.

AskCHIS is an online data query system and a resource for data about California, its people, and their health. There are many other online data query systems, and you will find that the concepts and skills learned with AskCHIS translate to other data sets and query systems.

CHIS Data Highlights:

- CHIS randomly selected up to 50,000 households, drawn from every county in California, for its random-digit dial telephone survey;
- CHIS has provided data to the public in a two year data cycle since 2001. Beginning in 2011, CHIS now conducts continuous sampling, gathering data more slowly over a longer period of time. Continuous data collection will allow CHIS to generate one-year estimates in addition to the two year data which will be released during the summer every other year as previously scheduled. These new one-year estimates will be available at the state level and for some heavily populated areas within the state, but not for smaller geographic areas. Because of the smaller sample size of the one-year data, the Center will release these estimates through select Center products, including publications as well as through the Center's easy-to-use Health Profiles fact sheets. There will be no PUF files available for one-year data nor will one-year data be available on [AskCHIS](http://askchis.org). For more information on the continuous sampling method, visit: <http://healthpolicy.ucla.edu/newsroom/press-releases/pages/details.aspx?NewsID=121>.
- CHIS over-samples certain population groups. Over-sampling is a technique that helps researchers make sure small population groups are represented in the sample. For example, in CHIS 2001, American Indian/Alaska Natives (AIAN) were over-sampled (raising their total sample size from 440 to 800) in order to examine important differences between AIAN living in urban and rural areas.
- CHIS is conducted in several different languages. Surveys were conducted in 5 languages (English, Spanish, Chinese, Korean, and Vietnamese);
- CHIS interviews one adult in each household. In households where there is a child, parents are asked to provide information about one child. In households where there is an adolescent, the adolescent is asked to participate by being interviewed.
- CHIS interviewed over 50,000 adults and 4,000 adolescents (ages 12 to 17). In addition, over 9,000 parents provided information about their children 11 years old and younger.
- Researchers applied a statistical formula to the data that was collected in order to create **population estimates**. The results you get using AskCHIS are population estimates, based on the geography and population characteristics (such as age and gender) that you select. For example, based on the number of adults in the survey who reported having diabetes, we were able to make an estimate for how many adults in California have diabetes.

CHIS data can be accessed in multiple ways:

- AskCHIS- the online data query system that is the focus of this workshop
- Data Access Center—located at the UCLA Center for Health Policy Research
- Public Use Files—available online via the AskCHIS website
- Requests can be made to UCLA Center for Health Policy Research for larger data analysis projects, please email Brandon Traudt: btraudt@ucla.edu

Using AskCHIS

1. Find the AskCHIS Website

Go to the CHIS website: <http://www.askchis.com/> and the AskCHIS login page seen in the below screenshot will immediately be visible.

AskCHIS gives you the answers

AskCHIS is a tool that allows you to quickly search for health statistics on your county, region and state. AskCHIS draws upon the responses of more than 50,000 Californians interviewed by The California Health Interview Survey (CHIS) - the largest state health survey in the United States.

AskCHIS is:

- Easy to use
- Quick
- Free
- Authoritative


You can see your results on-screen or export to an Excel file.

Thousands of health experts, policymakers, journalists and ordinary people use AskCHIS to quickly get data for grant proposals, needs assessments, research, news reporting, and policy-making.

AskCHIS login
 Username:
 Password:

[Forgot Your Password?](#)

Not registered?
[Create an account](#)


 The California Endowment

AskCHIS is supported by a grant from The California Endowment

2. Register to use AskCHIS (or log in, if you've already registered)

The first time you use AskCHIS, you'll need to register yourself as a user. You need an email address to do this. Click on the green box that says, "Not registered? Create an account" (See the arrow ← in the screen shot below). Fill in the registration form, and be sure to create a username and password that are easily remembered (if you forget, an email reminder can be sent). The information you provide is kept in strictest confidentiality and only collected to help identify who uses the site and what CHIS information/sections are used. The information gathered are used to determine what enhancements are needed to better serve our users. After you register once, you can use your username and password to log in each time your access AskCHIS.

The login screen is the first step to using AskCHIS:

AskCHIS

AskCHIS gives you the answers

AskCHIS is a tool that allows you to quickly search for health statistics on your county, region and state. AskCHIS draws upon the responses of more than 50,000 Californians interviewed by The California Health Interview Survey (CHIS) - the largest state health survey in the United States.

AskCHIS is:

- Easy to use
- Quick
- Free
- Authoritative

You can see your results on-screen or export to an Excel file.

Thousands of health experts, policymakers, journalists and ordinary people use AskCHIS to quickly get data for grant proposals, needs assessments, research, news reporting, and policy-making.

The California Endowment

AskCHIS is supported by a grant from The California Endowment

AskCHIS login

Username:

Password:

Submit

[Forgot Your Password?](#)

Not registered? [Create an account](#)

3. Find Data Using AskCHIS

AskCHIS allows you to quickly and easily get data from the California Health Interview Survey. There are four steps: Geographic Area, Main Topic, Compare By, and Population before you can get results.

AskCHIS

GEOGRAPHIC AREA | MAIN TOPIC | COMPARE BY | POPULATION | GET RESULTS


Search Query Summary: GEOGRAPHIC AREA:

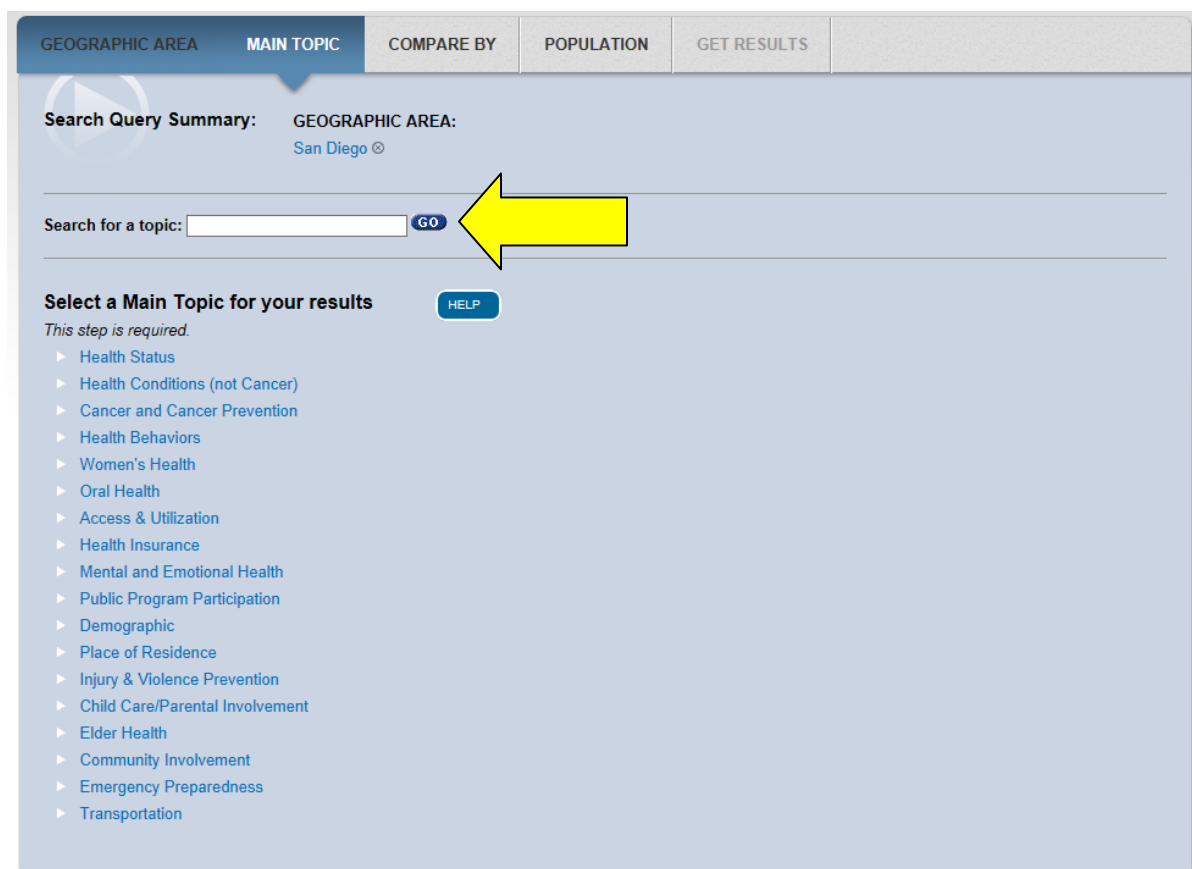
3.1. Select a Geographic Area for Your Results

First select your geographic area of interest (the entire state of California, a county or group of counties). When selecting specific geographic areas



other than the entire state, you will need to then click **Update Geographic Area**.

3.2 Select a Main Topic for Your Results

There are a number of basic topic categories provided on this screen. The easiest way to search for variables to select for your query is a Keyword Search. Type a key word in the box (see the arrow  in the screen shot below). For practice, type "asthma" and click on **GO**.



The screenshot displays the 'MAIN TOPIC' tab in a web application. At the top, there are navigation tabs: 'GEOGRAPHIC AREA', 'MAIN TOPIC', 'COMPARE BY', 'POPULATION', and 'GET RESULTS'. Below these, a 'Search Query Summary' shows 'GEOGRAPHIC AREA: San Diego'. A search box labeled 'Search for a topic:' contains a text input field and a 'GO' button, with a yellow arrow pointing to the 'GO' button. Below the search box, the text 'Select a Main Topic for your results' is followed by a 'HELP' button. A note states 'This step is required.' Below this, a list of topic categories is shown, each with a right-pointing arrow: Health Status, Health Conditions (not Cancer), Cancer and Cancer Prevention, Health Behaviors, Women's Health, Oral Health, Access & Utilization, Health Insurance, Mental and Emotional Health, Public Program Participation, Demographic, Place of Residence, Injury & Violence Prevention, Child Care/Parental Involvement, Elder Health, Community Involvement, Emergency Preparedness, and Transportation.

Once you type in a keyword, such as "asthma", your search results will look similar to this. Notice the **Select** button on the right-hand side and the Info button located just behind the Main Topic name (see the arrow  and arrow  in the screen shot below).

GEOGRAPHIC AREA **MAIN TOPIC** COMPARE BY POPULATION GET RESULTS

Search Query Summary: GEOGRAPHIC AREA:
San Diego ☒

Search for a topic: **GO**

Select a Main Topic for your results **HELP**
This step is required.
[See topics by category](#)

Search results for "asthma"

Ever diagnosed with asthma?   **SELECT ▶**
Most recent data year: 2009

Age groups: **Child Teen Adult**
listed under Health Conditions (not Cancer) -> Asthma

Currently have asthma after prior diagnosis by doctor? **SELECT ▶**
Most recent data year: 2009

Click on **Info** to find out more about a variable. A new window with information on the variable will open:

[x] CLOSE

Ever diagnosed with asthma
Years: 2001, 2003, 2005, 2007, 2009

Age groups included:
Child **Yes**
Teen **Yes**
Adult **Yes**

Results are displayed for the following levels:
Has asthma
Does not have asthma

2001
Respondents were asked: "Has a doctor ever told you that you have asthma?"
Adult Survey: AB17
Child Survey: CA12
Adolescent Survey: TB5

This variable is not asked of everyone:
Asked of all respondents 1 year of age and older

[x] CLOSE

SELECT ▶

Clicking the **Select** button will add this topic to your query. You will immediately be returned back to the query design page and will be prompted to choose a comparison demographic or trait. The Main Topic you chose will be displayed at the top of the page

The screenshot shows a query design interface with a navigation bar at the top containing five tabs: 'GEOGRAPHIC AREA', 'MAIN TOPIC', 'COMPARE BY optional', 'POPULATION', and 'GET RESULTS'. The 'COMPARE BY optional' tab is currently selected. Below the navigation bar, the 'Search Query Summary' section displays the following information:

- GEOGRAPHIC AREA:** San Diego ⊗
- MAIN TOPIC:** Ever diagnosed with asthma ⓘ ⓘ

At the bottom of the summary section, there is a search input field labeled 'Search for a topic:' followed by a 'GO' button.

3.3 Select a Compare By Group

Now we have the option to select a demographic topic or particular trait to compare by. If we select **a Compare By** group, our results generated by the query will include a two-way table with detail based on this comparison (two-way tables explained in detail in **Part 3: Executing Queries and Retrieving Data** on page 46). If we do not want to group the example based on any trait or indicator we can skip this section by simply clicking on the next section, **Population**. Adding a **Compare By** group is a way you can filter your data into categories. For our example, let us say we would like to categorize our data by age. We can do a keyword search or we can click on the categories as seen below. First click on the Demographics heading and then age name (see the arrow → and arrow ← in the screen shot below).

GEOGRAPHIC AREA **MAIN TOPIC** **COMPARE BY** optional **POPULATION** **GET RESULTS**

Search Query Summary: **GEOGRAPHIC AREA:**
 San Diego ⊗
MAIN TOPIC:
 Ever diagnosed with asthma ? ⊗

Search for a topic: **GO**

Compare By demographic topic/particular trait **Common Compare By topics include:** Currently insured **SELECT ▶**
 HELP Gender **SELECT ▶**

- ▶ Health Status
- ▶ Health Conditions (not Cancer)
- ▶ Cancer and Cancer Prevention
- ▶ Health Behaviors
- ▶ Women's Health
- ▶ Oral Health
- ▶ Access & Utilization
- ▶ Health Insurance
- ▶ Mental and Emotional Health
- ▶ Public Program Participation
- ▶ **Demographic**
 - Age & Gender ←
 - Language

This then opens up an additional section allowing you to choose from a variety of ways to categorize by age. To select your preferred Topic to Compare by simply click the select button (see the arrow → in the screen shot below).

Compare By demographic topic/particular trait **Common Compare By topics include:** Currently insured **SELECT ▶**
 HELP Gender **SELECT ▶**

- ▶ Health Status
- ▶ Health Conditions (not Cancer)
- ▶ Cancer and Cancer Prevention
- ▶ Health Behaviors
- ▶ Women's Health
- ▶ Oral Health
- ▶ Access & Utilization
- ▶ Health Insurance
- ▶ Mental and Emotional Health
- ▶ Public Program Participation
- ▶ **Demographic**
 - Age & Gender**
 - Language
 - Race/Ethnicity
 - Family & Marital Status
 - Education
 - Employment
 - Income & Poverty
 - Residency in US
 - Sexual Orientation
 - Military Service

Select a topic to Compare By

Age groups - 4 categorical levels ? **SELECT ▶**

Multiple years up to: 2009

Age groups: Child Teen Adult

Age groups - 9 categorical levels ? **SELECT ▶**

Multiple years up to: 2009

Age groups: Child Teen Adult

Gender ? **SELECT ▶**

Multiple years up to: 2009

Age groups: Child Teen Adult

Age groups - 19 categorical levels ? **SELECT ▶**

Multiple years up to: 2009

Age groups: Child Teen Adult

3.4 Select a Population for Your Results

In this section we can select results based on age, race, gender, and % of Federal Poverty limit (FPL). Now we select our population of interest. Let's limit the results to include only females. In the Gender box, select **Female** (see the arrow ← in the screen shot below).

GEOGRAPHIC AREA
MAIN TOPIC
COMPARE BY
POPULATION
optional
GET RESULTS

Search Query Summary:

GEOGRAPHIC AREA: San Diego ⊗

MAIN TOPIC: Ever diagnosed with asthma ⊗ ⊗

COMPARE BY: Age groups - 4 categorical levels ⊗ ⊗

Limit the Population included in your results HELP

This step is optional. You may select any combination of the below settings.

Age

Include all ages

Include results for:

 to years of age.

- The age range includes through (i.e., not up to) the years specified.
- For infants under 6 months, enter 0 years.
- You may leave either box blank. [More tips](#)

Race*

Include all races

Latino

White (non-latino)

African American (non-latino)

American-Indian/Alaska Native (non-latino)

Asian (non-latino)

Native Hawaiian/Pacific Islander (non-latino)

Two or More Races (non-latino)

*Based on the Office of Management and Budget (OMB) definitions. You may find "RACE - UCLA definition" in **Compare By (Demographics)**. [See more about Race \(PDF\)](#).

Gender

Include both genders

Male

Female ←

Federal Poverty Level (FPL)

Include all FPL

Include results for:

 to

The left box can be 0% or any number above 99%.
The right box can be any number 99% to 500%, or leave it blank for no upper limit
[See more information about Federal Poverty Level guidelines](#)

Accessing Data from the AskCHIS Online Data Query System

Part 2: Developing the Data Query

Learning Objectives:

- Understand the type of data AskCHIS can provide
- Assess the availability and appropriateness of data in AskCHIS
- Prepare the data query for AskCHIS

What data can AskCHIS provide?

Surveys such as CHIS allow data users to describe the population numerically, including frequencies of their characteristics, certain health-related behaviors, and other aspects of their relationships with institutions such as health care and insurance systems, the labor market, or public benefits (e.g., Aid for Families with Dependent Children or food stamp programs).

Let's take a look at the questions you provided in the earlier session, during the section, "Determining the Data You Need." These questions should be quantifiable, or measurable and answerable by **statistics, frequencies, averages, and counts**.

It is important to look for and consult the technical documentation that often accompanies an online data source before conducting any query. Consulting these resources can save time, resources, and frustration (especially if data that you are looking for is not available).

Resources to assess the availability and relevance of data in AskCHIS to your data needs include:


1. The AskCHIS topic fact sheet
Displays the general topic items and populations for whom data was collected (i.e. adults, adolescents, and/or children).
<http://healthpolicy.ucla.edu/chis/design/Pages/survey-topics.aspx>
2. The data dictionary (also known as the codebook)
Describes how questions were asked, the response categories available, and how variables were created.
<http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx>

3. The questionnaire
Provides the actual questions asked of CHIS survey respondents. Available in Child, Adolescent, and Adult versions.
<http://healthpolicy.ucla.edu/chis/design/Pages/questionnaires.aspx>
4. The AskCHIS Keyword Search tool (online)
This tool allows you to type in the term(s) of your choice and search for variables related to or containing that term. You must be logged in to AskCHIS to use this tool.

Defining The Data Query & Population and Preparing the Data Query

- Refer to the AskCHIS topic fact sheet (visit: www.chis.ucla.edu/topics.html) and look for the subject heading where your data of interest may be located.
- In **AskCHIS**, go to the **Select a MAIN TOPIC** page. Use the keyword search tool to verify the variable(s) available in AskCHIS for that topic. Type a word in the box and click on **GO**.

The screenshot shows the AskCHIS web interface. At the top, there are tabs for 'GEOGRAPHIC AREA', 'MAIN TOPIC', 'COMPARE BY', 'POPULATION', and 'GET RESULTS'. Below the tabs, there is a 'Search Query Summary' section showing 'GEOGRAPHIC AREA: Entire state of California'. Below that is a search box with the text 'Search for a topic:' and a 'GO' button. The main section is titled 'Select a Main Topic for your results' and includes a 'HELP' button. A list of topics is displayed, with a yellow arrow pointing to 'Access & Utilization'. The topics listed are: Health Status, Health Conditions (not Cancer), Cancer and Cancer Prevention, Health Behaviors, Women's Health, Oral Health, Access & Utilization, Health Insurance, Mental and Emotional Health, Public Program Participation, Demographic, Place of Residence, Injury & Violence Prevention, Child Care/Parental Involvement, Elder Health, Community Involvement, Emergency Preparedness, and Transportation.

- Some people like finding the variables from a list of topic categories. You can do this by clicking on any of the categories listed on the left side of the page (See the arrow  on the left in the screen shot above.)
- Record the variable(s) of interest to you on Worksheet 1. (An additional copy of the worksheet can be found in Appendix C of this workbook.)

You may not find exactly the data you are looking for or need. If this is the case, you have two options:

1. Refine your data query so that it can be answered by the type of data available in AskCHIS in that general subject area; or
2. Identify a new data query.

Accessing Data from the AskCHIS Online Data Query System

Part 3: Executing Queries & Retrieving Data

Learning Objectives:

- Navigate the AskCHIS internet-based data query system to execute queries.
- Access data using the AskCHIS query system.
- Understand what 1--way and 2--way tables are and how to interpret them.
- Develop 1-way and 2--way tables.
- Assess the reliability of results obtained

Review of the key steps for using AskCHIS:

- Go to the CHIS webpage, <http://www.askchis.com> and click on AskCHIS
- Register (or log in, if you've already registered)
- Select a geographic area for your results
- Select main topic for your results
- Select a population for your results
- Review the results in the data table


GETTING RESULTS!

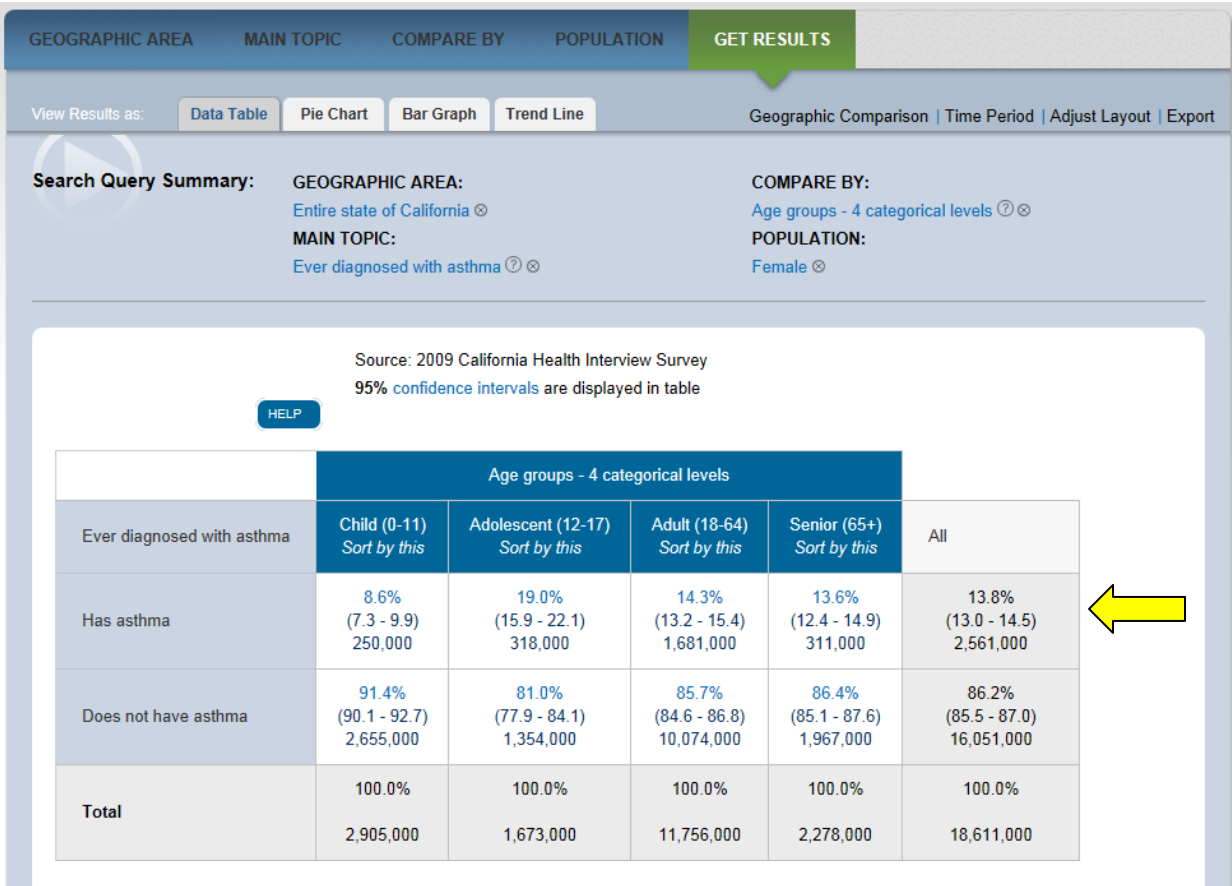
To see your results, click **Get Your Results** at the top of this same page (see the arrow ← in the screen shot below).

The screenshot shows the AskCHIS interface with the following elements:

- Navigation Bar:** GEOGRAPHIC AREA, MAIN TOPIC, COMPARE BY, POPULATION (optional), and GET RESULTS (highlighted with a yellow arrow).
- Search Query Summary:**
 - GEOGRAPHIC AREA: Entire state of California
 - MAIN TOPIC: Ever diagnosed with asthma
 - COMPARE BY: Age groups - 4 categorical levels
- Limit the Population included in your results:** (This step is optional. You may select any combination of the below settings.)
 - Age:**
 - Include all ages
 - Include results for: [] to [] years of age.
 - The age range includes through (i.e., not up to) the years specified.
 - For infants under 6 months, enter 0 years.
 - You may leave either box blank.
 - [More tips](#)
 - Race***
 - Include all races
 - Latino
 - White (non-latino)
 - African American (non-latino)
 - American-Indian/Alaska Native (non-latino)
 - Asian (non-latino)
 - Native Hawaiian/Pacific Islander (non-latino)
 - Two or More Races (non-latino)
 - *Based on the Office of Management and Budget (OMB) definitions. You may find "RACE - UCLA definition" in Compare By (Demographics). [See more about Race \(PDF\)](#).
 - Gender:**
 - Include both genders
 - Male
 - Female
 - Federal Poverty Level (FPL):**
 - Include all FPL
 - Include results for: [] to []
 - The left box can be 0% or any number above 99%.
 - The right box can be any number 99% to 500%, or leave it blank for no upper limit
 - [See more information about Federal Poverty Level guidelines](#)
- Add more options above:**

To further limit your results to a certain group of people, you can add more boxes above. Type in a keyword that describes a trait for the group of people that interests you

A result page will generate with a two-way table displayed. What percentage of females in California have been diagnosed with asthma? (See the arrow  in the screen shot below).



Source: 2009 California Health Interview Survey
95% confidence intervals are displayed in table

[HELP](#)

Ever diagnosed with asthma	Age groups - 4 categorical levels				All
	Child (0-11) <i>Sort by this</i>	Adolescent (12-17) <i>Sort by this</i>	Adult (18-64) <i>Sort by this</i>	Senior (65+) <i>Sort by this</i>	
Has asthma	8.6% (7.3 - 9.9) 250,000	19.0% (15.9 - 22.1) 318,000	14.3% (13.2 - 15.4) 1,681,000	13.6% (12.4 - 14.9) 311,000	13.8% (13.0 - 14.5) 2,561,000
Does not have asthma	91.4% (90.1 - 92.7) 2,655,000	81.0% (77.9 - 84.1) 1,354,000	85.7% (84.6 - 86.8) 10,074,000	86.4% (85.1 - 87.6) 1,967,000	86.2% (85.5 - 87.0) 16,051,000
Total	100.0% 2,905,000	100.0% 1,673,000	100.0% 11,756,000	100.0% 2,278,000	100.0% 18,611,000

Introduction to One-Way Tables

What are they?

- The term '**one-way table**' refers to a table that organizes the data in a simple manner.
- The table provides the answer to a data query that only focuses on one health topic or characteristic.
- These are also known as univariate tables.

What do they look like?

- A one-way table has several components:
 - One variable (known in AskCHIS as **Main Topic**), and

- The demographic factors that define the population of interest (known in AskCHIS as **Population**).

- Here is an example of a query that produces a one-way table:
 - *How many adults (or, what percentage) in California have health insurance?*
- This is a one-way analysis because only one variable (health insurance) is required to obtain the statistic, that is, percentage of California adults with health insurance.

When are they used?

- One-way tables are used when a person is interested in answering a broad question, such as: *How many people have a certain characteristic?*

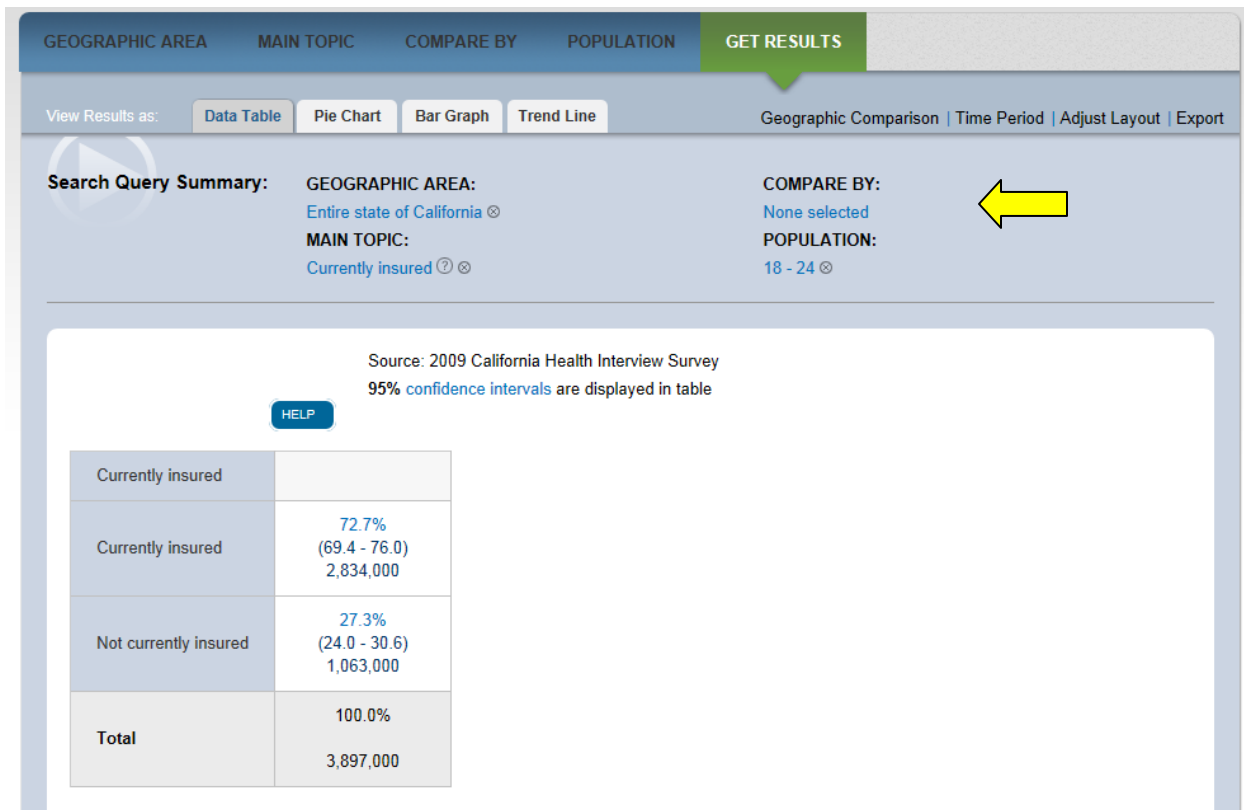
What types of data do they provide?

- In AskCHIS, one-way data tables provide percentages, population estimates, and confidence intervals.

Now we'll practice a simple query, using a previously created data query as a guide. Our results will be a one-way table. As the trainer leads you through the steps, you can make selections for Geography, Main Topic, and Population that relate to your work and interests. [See Example One on the next page.] After this exercise, you will have additional time to practice queries for one-way tables.

Discussion Questions:

- Try to interpret your findings.
 - What do these results mean?
 - What population(s) were included in the results? Describe the population by:
 - Geographic region,
 - Age,
 - Gender,
 - Income, and/or
 - Race / ethnicity (as needed)



Please note on the results page we can clearly see all selected criteria listed in the top portion of the screen (See the arrow ← in the screen shot above).

Interpretation of Data Results - Example 1

- In California in 2009, 72.7% of adults ages 18-64 had health insurance.
- In California in 2009, 27.3% of adults ages 18-64 did not have health insurance.

Introduction to Two-Way Tables

What are they?

- Two-way tables build on the analysis conducted earlier. They answer more complex questions, since two variables (rather than one) are being analyzed.
- Two-way tables are also known as bivariate tables.

What do they look like?

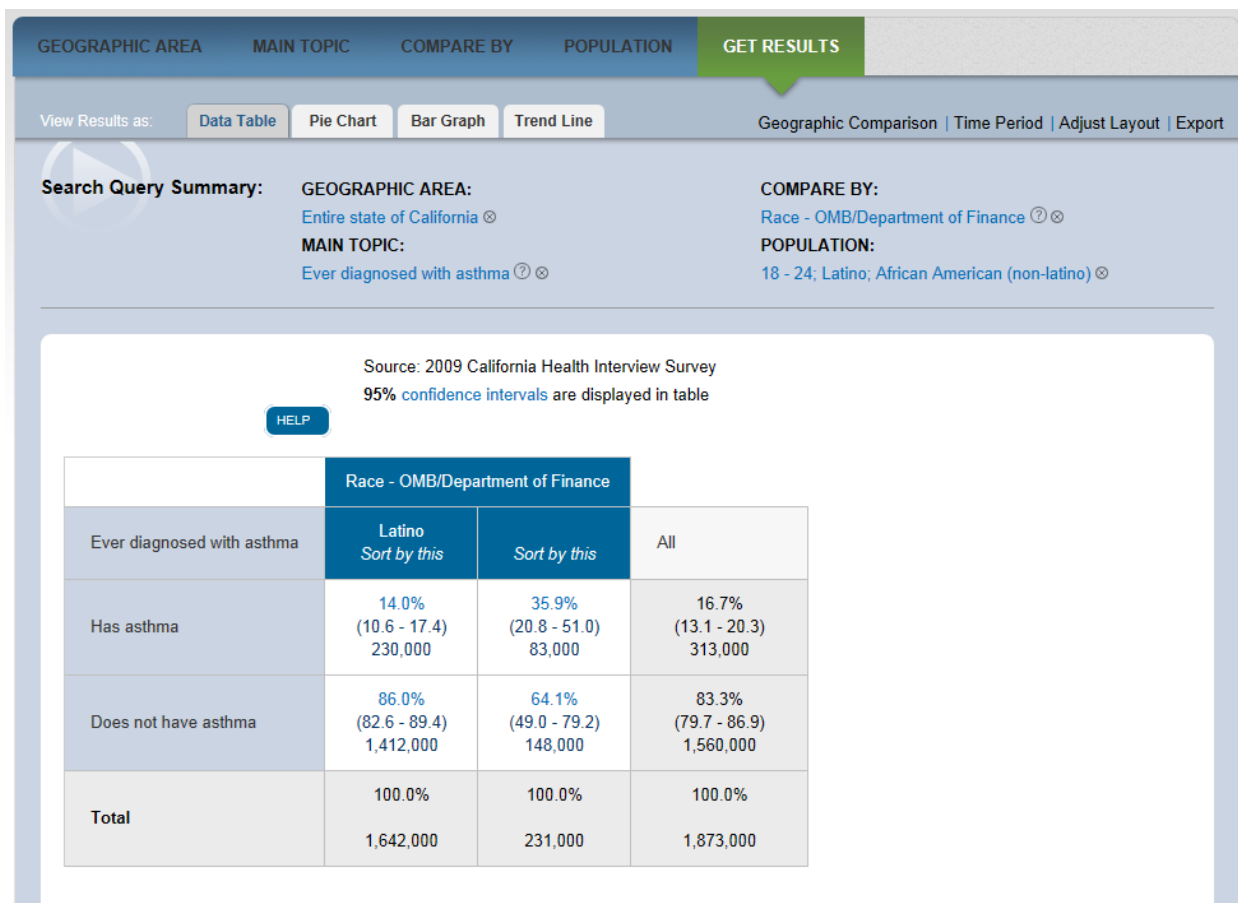
- A two-way query is made up of two variables (a **Main Topic** and a **Compare By** variable that you will use for comparison), plus the demographic factors that define the population.
- Two-way tables differ from one-way tables because they add a second variable to the analysis. When using AskCHIS the Compare by option is provided making it easy to create a two-way table.

Here are a few examples:

- We might ask: *Does the percent of uninsured adults differ between those with diabetes and those without?*
- This query allows us to explore two AskCHIS variables at the same time. The two-way results table will allow us to look at the distribution of uninsurance among diabetic and non-diabetic adults.
- We might ask: *Does the percentage of adults diagnosed with heart disease differ between normal weight and obese adults?*
- This query allows us to explore two AskCHIS variables at the same time. The two-way results table will allow us to look at the distribution of heart disease diagnosis across weight categories.

When are they used?

Two-way tables are used when a person is interested in finding out how one AskCHIS variable (such as current insurance) is distributed across the levels of a second variable (such as income).



Interpretation of Data Results - Example 2

- In California in 2009, 14.0% of Latinos in California aged 18-24 had ever been diagnosed with asthma.
- In California in 2009, 35.9% of African Americans in California aged 18-24 had ever been diagnosed with asthma.

Empty Cells and Unstable Estimates:

In some queries, especially those limited to a small population subset, AskCHIS may either not report (through empty cells) or may report an unstable estimate (marked by a red star).

- Unstable estimates, characterized by Asterisks (*) in AskCHIS are of concern to data users because they reflect percentages based on a small sample size.
- Empty cells, characterized by a dash (--) in AskCHIS, are of concern because the data were not reportable due to either lack of responses to a certain question or small sample sizes (i.e., cell has a population estimate <500 individuals).
- If the query produces a table with values that are asterisked, the UCLA Center for Health Policy Research and CHIS Team do not recommend using these values for purposes of policy development, program planning, advocacy, etc.
- Consideration should be given to these issues in use of data. For example, you may not want to plan an entire program or secure funding based on a small sample that may not accurately reflect the behaviors, health conditions, or health services issues of the population. This may result in inappropriate allocation of limited resources.
- Nearly all query systems should give an indicator of data that are unusable, not available because of sample size issues, or unstable.
- **To avoid unstable estimates or missing data in your results, do not limit the population too much.** Try removing or changing the demographic variables that define your subset. In other words, broaden your query to include a larger population segment. For example, include a larger geographic area, larger age range, males and females, all race/ethnicities, etc. For more information on strategies for handling unstable estimates please refer to Worksheet #3 in Appendix C.

Refining your data query:


Now refine your query. If you had no unstable estimates the first time, practice limiting your population to see what happens to the results. If you had unstable estimates the first time, practice expanding your population subset.

Discussion Questions:

- Try to interpret your findings.
 - What do these results mean?
- Did you obtain unstable estimates or empty cells?
 - If so, what could you do to refine your query and produce usable data?

Additional Features for Results in AskCHIS:


There are several additional options for your results in AskCHIS including the ability to change the geographic area, time period, and table layout.

- **EXPORT TO EXCEL:** Data tables can be exported to Excel for further analysis. (see the arrow  in the screen shot below).



The screenshot shows the AskCHIS interface with the 'GET RESULTS' tab selected. The 'View Results as:' section includes 'Data Table', 'Pie Chart', 'Bar Graph', and 'Trend Line'. The 'Export' button is highlighted with a yellow arrow. The 'Search Query Summary' displays the following information:

Category	Value
GEOGRAPHIC AREA:	Entire state of California ⊗
MAIN TOPIC:	Ever diagnosed with asthma ? ⊗
COMPARE BY:	Race - OMB/Department of Finance ? ⊗
POPULATION:	18 - 24; Latino; African American (non-latino) ⊗

- **ADJUST TABLE LAYOUT:** Any changes to the layout of the table can be done by clicking on the Adjust Layout tab (see the arrow  in the screen shot below). This feature includes transposing and collapsing columns.



The screenshot shows the AskCHIS interface with the 'GET RESULTS' tab selected. The 'View Results as:' section includes 'Data Table', 'Pie Chart', 'Bar Graph', and 'Trend Line'. The 'Adjust Layout' button is highlighted with a yellow arrow. The 'Search Query Summary' displays the following information:

Category	Value
GEOGRAPHIC AREA:	Entire state of California ⊗
MAIN TOPIC:	Ever diagnosed with asthma ? ⊗
COMPARE BY:	Race - OMB/Department of Finance ? ⊗
POPULATION:	18 - 24; Latino; African American (non-latino) ⊗

- **CHANGING TIME FRAMES & TRENDING DATA:** One important feature is the ability to change in the time period to incorporate more data from various years or trend over time. We can select Time Period and then choose to compare two years or pool data together.



The screenshot shows the AskCHIS interface with the 'GET RESULTS' tab selected. The 'View Results as:' section includes 'Data Table', 'Pie Chart', 'Bar Graph', and 'Trend Line'. The 'Time Period' button is highlighted with a yellow arrow. The 'Search Query Summary' displays the following information:

Category	Value
GEOGRAPHIC AREA:	Entire state of California ⊗
MAIN TOPIC:	Ever diagnosed with asthma ? ⊗
COMPARE BY:	Race - OMB/Department of Finance ? ⊗
POPULATION:	18 - 24; Latino; African American (non-latino) ⊗

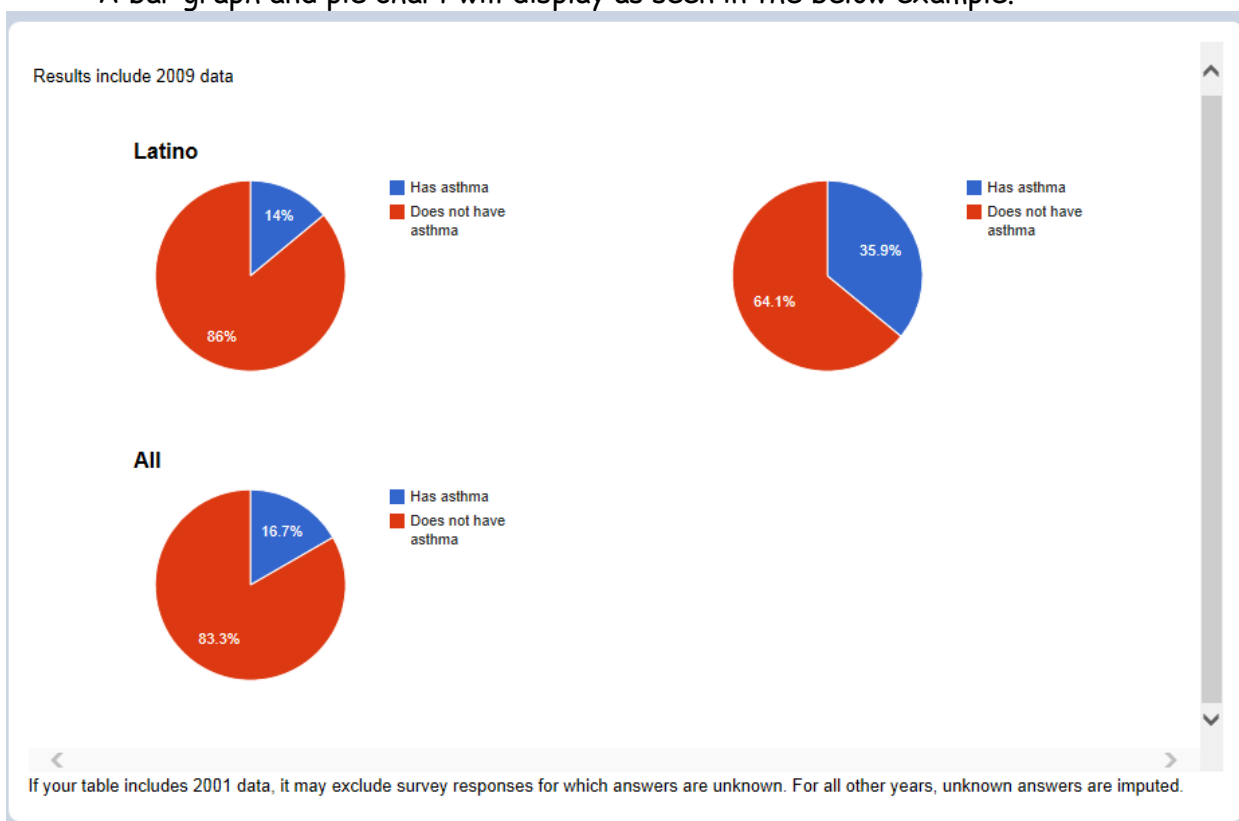
If we would like to see the # of Latino and African American individuals diagnosed with asthma trended with data from all the available CHIS surveys, we simply click the Trend Line icon (See the arrow ← in the screen shot below).

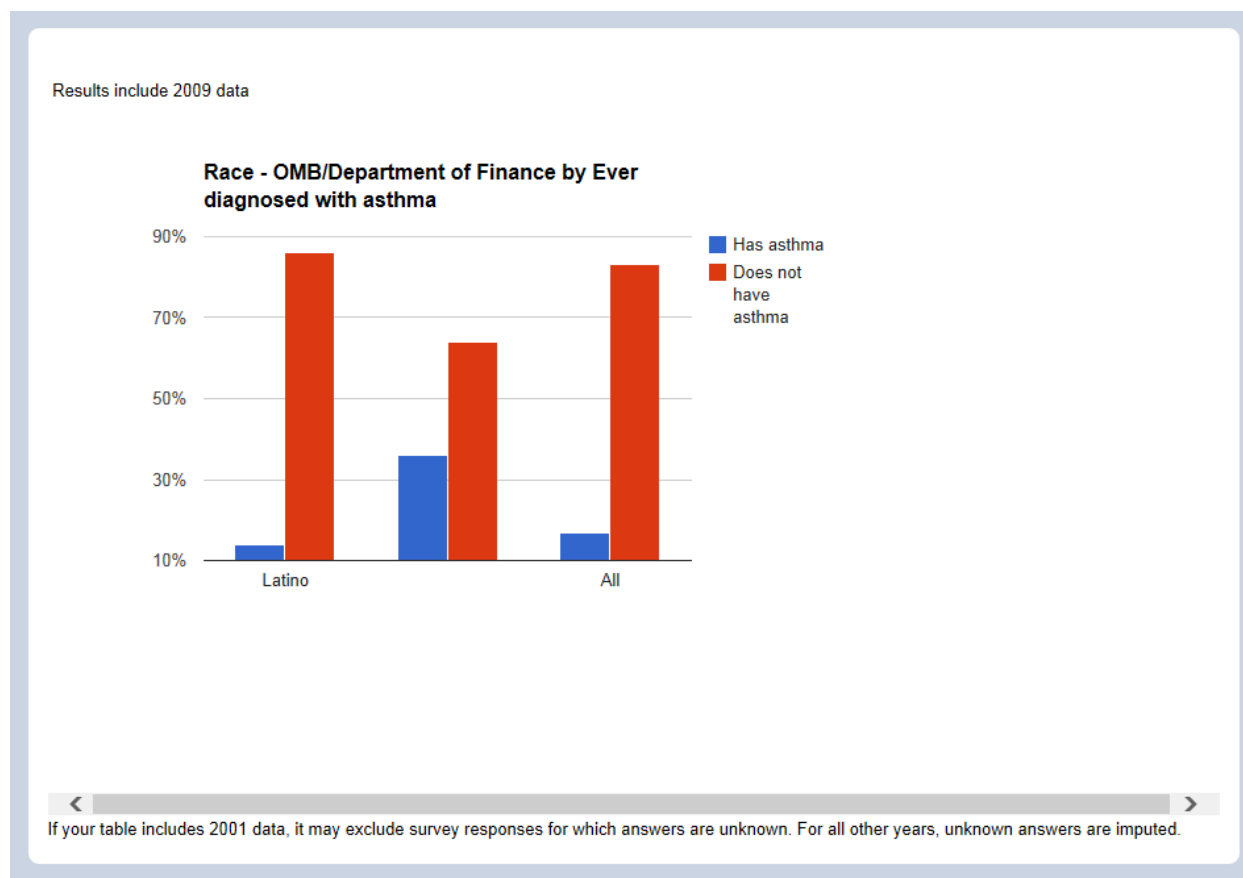
The screenshot shows the top navigation bar with tabs for GEOGRAPHIC AREA, MAIN TOPIC, COMPARE BY, POPULATION, and GET RESULTS. Below this is a 'View Results as:' section with buttons for Data Table, Pie Chart, Bar Graph, and Trend Line. A yellow arrow points to the Trend Line button. To the right are links for Geographic Comparison, Time Period, Adjust Layout, and Export. Below the navigation is a 'Search Query Summary' section with the following details:

GEOGRAPHIC AREA: Entire state of California ⊗	COMPARE BY: Race - OMB/Department of Finance ? ⊗
MAIN TOPIC: Ever diagnosed with asthma ? ⊗	POPULATION: 18 - 24; Latino; African American (non-latino) ⊗

- **BAR GRAPHS/VISUALS:** You can view your results in graphical form by clicking on Pie Chart and Bar Graph.

A bar graph and pie chart will display as seen in the below example.





Refining your data query:

Now refine your query. Explore your options for narrowing your population in AskCHIS. In the query above, for example, you can narrow the population to include only adults who live in a household where someone smokes. Do the asthma rates change when you refine the query?

Explore your options for comparison variables. In the query above, we compared by race. What do you find when you change the comparison variable to number of doctor visits in the last 12 months?

Now work on refining your own query. AskCHIS will prompt you to make a change if the variables selected for main topic, comparison variable, and characteristics narrowing the population are not ALL available for the same year (2001, 2003 or 2005).

Discussion Questions:

- Try to interpret your findings.
 - What do these results mean?

- Did you obtain unstable estimates or empty cells?
- If so, what could you do to refine your query and produce usable data?

Summary & Discussion of the AskCHIS Data Query Process

- Review the major steps involved in developing queries and obtaining 1 and 2 tables.
- Handouts and worksheets are in your packets as references for self-paced analyses.
- Support is available from Health DATA following the workshop if you need it.

Accessing Data from the AskCHIS Online Data Query System

Part 4: Interpreting and Presenting AskCHIS Data Findings

Learning Objectives:

- Understand how transposing items in the table changes the results.
- Identify whether the results answer the question you started with
- Understand how to localize data.
- Write out key message statements and interpretation of findings

Review:

- What is the purpose of a **one-way** table?
 - One-way tables describe a single behavior or characteristic within a population of interest.
- What is the purpose of a **two-way** table?
 - Two-way tables describe how a behavior or characteristic is distributed among ranges or levels of a second characteristic.

Interpreting and Applying Data Results:

- Using the data results obtained from your AskCHIS query, follow the worksheet to identify your audience, the goal of the analysis, interpretation of the results, and your key message based on findings.
- Those of you who have participated in other Health DATA workshops may consider how these results may be applicable to media advocacy or to a community assessment strategy.

Identifying Solutions for Stabilizing Estimates in Small Populations:

“Localizing” data:

Sometimes data is limited or unavailable for the population you need. In these cases, you can *localize data*, or take existing data and apply it to your population of interest. These methods include:

- Making your own estimate,
- Painting a picture, and
- Asking a researcher.

Make Your Own Estimate

1. Obtain data from a CREDIBLE source that resembles/approximates the data you need.
2. 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.
 - Do the differences affect the applicability of the data to your constituents? If not, you can probably make your own estimate.
3. Utilizing several data sources, you then, "piece together" data in order to arrive at an estimate for your target population.
 - You can do this, but remember to protect your credibility when you create an estimate.
 - 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.

Paint a Picture

- Is it possible to piece together data from several sources to illustrate your point?
- Do you have personal information, anecdotal or statistical, that can complement this information?
- Might it be possible to "paint a picture" of the problem using a combination of personal stories and statistics from several sources?

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.

Presenting AskCHIS Data Using Graphs:

How you present data can be almost as important as understanding and using it. Your audience must be able to understand what you are saying. Visual aids such as graphs and charts are some of the most widely used and effective ways of reinforcing the audience's understanding of the information.

The presentation of data depends on audience and application of data. There are numerous ways of presenting the results of your data queries:

- Numerical (data tables)
- Text or narrative
- Graphical

Graphical depictions include:

Pie Graph

- Use a pie graph when you have simple percentages and the "slices" of the pie are not too numerous.
- Ideal for depicting the size of each part as a percentage of the whole.
- Dividing the pie graph into too many "slices" can lead to confusion.
- Best when visual is displayed in color, difficult to discern between "slices" if using grayscale to represent different groups.

Bar Graph

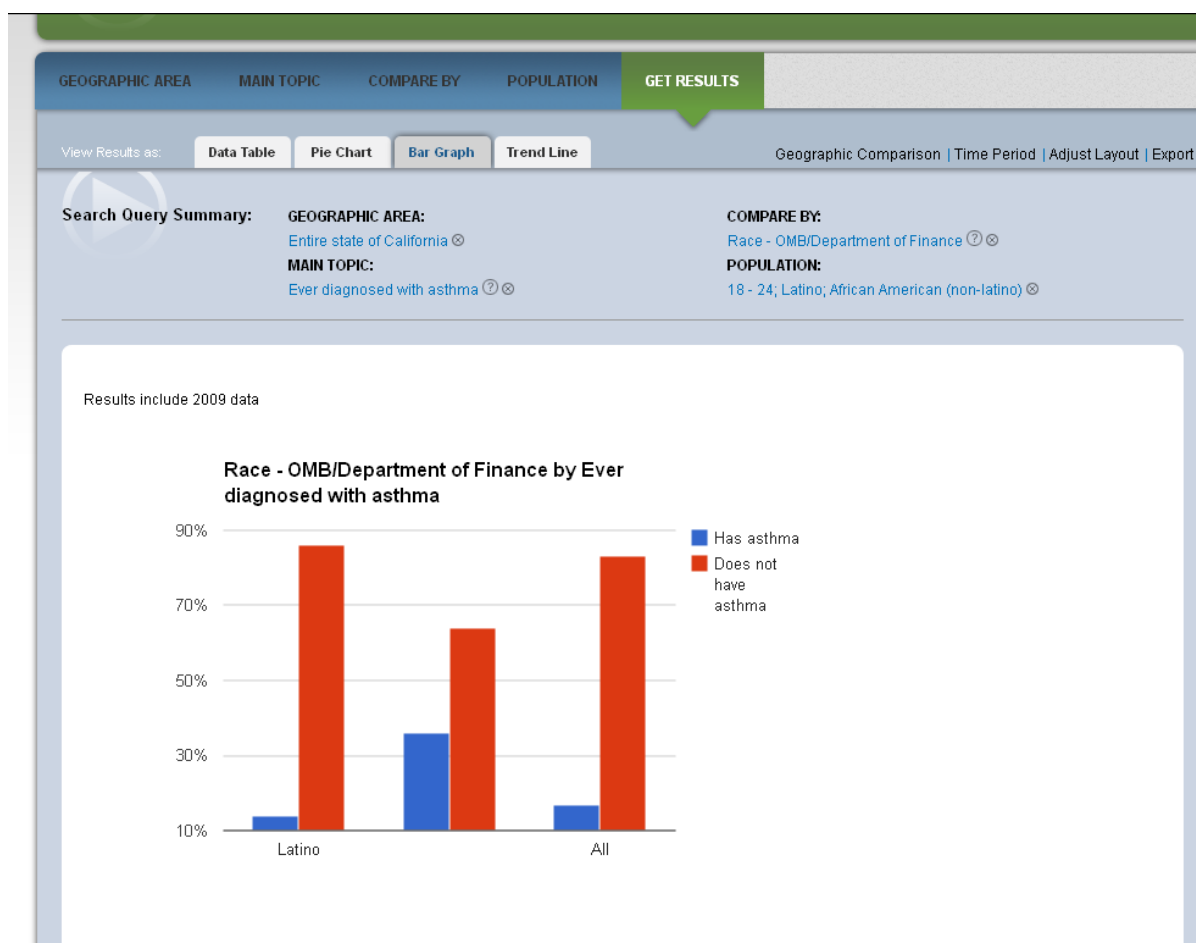
- Good for comparing quantities - simple bar lines are easy to read and compare.
- Avoid comparing things that are on different scales. Uneven scales can lead to confusion for your audience.

Chart/Table

- Suitable for providing simple numeric information
- Best used for side-by-side comparison of numbers and data for various variables or groups.

Creating and Saving Pie and Bar Graphs with AskCHIS:

1. Go back to the simple query we did in Example One (use the query you saved).
2. Click on **Get Results**.
3. To see a pie graph, click on **Pie graph**.
4. To see a bar graph, click on **Bar graph**.
5. The easiest way to save the graphs is to move the mouse arrow over Export. Choose one of the 3 options.
6. Is a pie graph or bar graph is better for your results? Why?
7. Repeat steps 1-5 for the more complex query we did in Example Three (use the query you saved).



Appendices

Appendix A: Health DATA Toolkit
Combining Quantitative & Qualitative Data
In Focus: Confidence Interval
What to Do When the Data You Need is Not Available

Appendix B: AskCHIS Worksheets & Examples

Appendix C: Data Resources

APPENDIX A: HEALTH DATA TOOLKIT
COMBINING QUANTITATIVE & QUALITATIVE DATA
IN FOCUS: CONFIDENCE INTERVAL
WHAT TO DO WHEN THE DATA YOU NEED IS NOT
AVAILABLE

Health DATA Toolkit

SECTION I: TYPES OF DATA

A. Qualitative Data: Data that is usually measured and expressed in the form of words, concepts, themes, or categories rather than numbers. Qualitative data are often used to gain a more in-depth understanding of a particular incident or phenomenon - they answer how or why something is occurring.

Qualitative techniques include, but are not limited to:

- **Observation:** Noting and recording information about a subject or group of subjects. The observer gathers information by watching and listening to the subject or subjects in their natural environment, and strives not to intrude or disrupt the action.
- **Ethnography:** The study and subsequent recording of information about human culture.
- **Case study:** A study based on an intensive observation of one (or a few) cases or examples, such as organizations or events.
- **Open-ended interview*:** Gathering information by speaking with an individual using questions that cannot be answered by merely "yes" or "no".
- **Focus group:** a group of individuals led through a structured discussion of a particular topic or event. Focus groups are often used to assess social needs, develop hypothesis and survey questions, investigate the meaning of survey results, and assess the range of opinions.

B. Quantitative Data: Data that is usually measured and expressed in the form of numbers or statistics and which usually answer the who, what, when and where questions of a research problem.

Quantitative techniques include, but are not limited to:

- **Census:** A complete enumeration of the population
- **Survey*:** A systematic way of collecting information from a defined population, usually by means of interviews or questionnaires administered to a sample of the population.
- **Questionnaire*:** A method of collecting data by asking participants identical questions about a particular issue or issues. Questions may be open-ended (the answer is completely left up to the respondent) or close-ended (where respondents are presented with a limited number of options to reply, such as yes/no, true/false or Likert-scale responses).

- **Close-ended interview***: Using a structured set of "yes" or "no" questions to gather information from an individual about a specific area of study.

Combining Quantitative and Qualitative Data

	Quantitative	Qualitative
Description	<ul style="list-style-type: none"> • Measured and expressed in the form of numbers or statistics • Also called numeric data • Can answer the who, what, when and where of an issue 	<ul style="list-style-type: none"> • Measured and expressed in the form of words, concepts or categories • Also called anecdotal data • Can answer the how or why • Can explore issues more in-depth
Data Collection Methods	<ul style="list-style-type: none"> • Secondary data • Surveys • Interviews 	<ul style="list-style-type: none"> • Observations • Focus groups • Surveys • Interviews
Benefits	<ul style="list-style-type: none"> • Can demonstrate cause and effect • Can "represent" a community by capturing the perspectives of many respondents • Usually easier to interpret 	<ul style="list-style-type: none"> • Richer, more in-depth information about the topic being studied • Can provide data from a respondent in their own words • Can collect new data and new ideas in a dynamic and unstructured way
Drawbacks	<ul style="list-style-type: none"> • Unable to provide rich, in-depth data • Cannot collect new ideas or responses; restricted to the data that has already been collected 	<ul style="list-style-type: none"> • Cannot demonstrate cause and effect • Usually not able to "represent" a community • Can be difficult to interpret

* NOTE: Interviews, Questionnaires, and Surveys can be quantitative, if utilizing a close-ended format, or they can be qualitative, if an open ended format is used.

SECTION II: GLOSSARY OF USEFUL DATA TERMS BY TOPIC

TOPIC A: METHODOLOGY - "HOW DID YOU GET THE DATA?"

The following terms are useful when describing *how* data were collected. The terms can apply to both qualitative and quantitative data.

Anecdote (or anecdotal data): A particular incident or fact of note used to illustrate a situation. Often, a descriptive story or information gained from day-to-day experiences, but not systematically.

Bias: Any factor that prevents research from providing a representative sample for the population being studied.

Cross-sectional: A cross-sectional research design is one in which all data are collected at one point in time. The purpose is to easily identify whether there is an association between two variables.

Estimate: A number drawn from a sample that is intended to describe the population from which the sample is drawn.

Likert-scale: A research design that utilizes scaled responses such as strongly agree, agree, neutral, disagree, and strongly disagree to elicit participant feelings about a research topic.

Longitudinal: A study in which data from the same population are gathered at two or more distinct points in time. This allows the researcher to identify trends and changes over time and more closely determine if there is a cause and effect association between two variables.

Methodology: A structured approach used to answer specific questions or test the hypothesis in a study. Can be called the "science of finding out."

Population: Entire group of individuals, who are logically related in some way, about which information is desired.

Random selection: A method of research that assigns every member of a group the same chance of being chosen to be a respondent. (Note: Choosing every 10th person from an alphabetized list is not random since some ethnicities will have names that cluster in a few letters, while others will be spread throughout; in contrast, picking names out of a hat is random.)

Random Assignment or randomization: Assignment of research subjects by chance into a study or survey. Randomization helps to eliminate potential bias.

Response rate: The proportion of subjects who were asked to participate in the study that completed the survey. Poor response rates may indicate some bias because those who chose to respond may be different in important ways from those who did not respond.

Sample: The subset of a population selected for study. Because it is often costly to interview a whole population (for example, the entire population of California), a sample of Californians are interviewed, who represent the whole population. They are usually selected based on the criteria (variables) essentially to answering the research question.

Triangulation: Determining the applicability of data gathered from a variety of sources, using differing research methods, to answer a particular research question.

Variable: Any characteristic of a research participant that can be expressed as a number. Variables are often in the form of attributes such as sex, age, employment status, income, etc.

TOPIC B: DATA INTERPRETATION - "WHAT DO THE NUMBERS MEAN?"

These terms are important when describing the *message* that the data contain. After data have been collected, these terms communicate what the data suggest or prove. These terms are used to interpret numerical data; therefore, they are used to describe quantitative data.

Average: the value that is intended to represent the general value, central tendency, of a set of unlike numbers. It is computed by adding the values of all the cases and dividing by the total number of cases. The term average is synonymous with the term "mean".

Cause: the reason something happens.

Confidence Interval: a confidence interval is a range around a measurement that conveys how precise the measurement is. See the "IN FOCUS Handout" for a comprehensive explanation.

Correlation: two events that change in concert with each other. A correlation does not mean cause and effect. For example, early in the AIDS epidemic it was noticed that the more often people used "poppers" (the stimulant amyl nitrate) the more likely they were to get AIDS. While these two were "correlated," it turned out that poppers were not the cause of AIDS, but were most commonly used by gay men during sex.

Incidence: The number of NEW cases during a period of time (e.g. the number of people newly diagnosed with HIV in the year); this number is useful to tell you about changes in the number of people affected by the condition.

Indicator: A measure that is closely associated with or a component of a condition you are studying. For example, mortality is a "health indicator" -- higher death rates are taken to mean that a population is less healthy even though death is a very crude way to measure "health."

Mean: See average.

Median: The value in a distribution of numbers that falls directly in the middle, such that 50% of the values lie below and 50% of the values lie above that value.

Mode: The most frequent value in a distribution. Also termed "probability average."

Percentage: A proportion where the denominator is expressed as 100; $\frac{1}{4}$ is 25/100, expressed as 25%. (See proportion)

Prevalence: The number of EXISTING cases at a point in time (e.g. the number of people living with HIV at this moment).

Proportion: The number of persons (or events) of interest divided by the total number of persons in the population (or events). If 100 persons have a disability in a community of 400, then $\frac{1}{4}$ are disabled.

Range: The true upper limit in a distribution minus the true lower limit. It is measure of variability.

Rate: Indicates the frequency of a given event (e.g., 100 births per 1,000 adults). It is a way of knowing the proportion of a population, possessing a particular variable, in order to compare areas or groups of different sizes.

Standardized rate: Populations often differ in characteristics related to the health issue under consideration, such as birth rates that vary with age. A younger population will have a higher birth rate than an older population. Standardizing portrays rates "as if" the characteristics of the two populations were the same (for birth and death rates it is often called age adjusted rates).

Statistical significance: The characteristic of an association that is not likely to be due to chance. In statistical terms, data are statistically significant if they meet certain criteria, often that the probability of error (p) is less than 5 out of 100 ($p < .05$).

IN FOCUS: Confidence Interval

Statistical Definition (www.cirem.org.uk/definitions.html):

"A confidence interval is an interval used to estimate the likely size of a population parameter. It gives an estimated range of values (calculated from a given set of sample data) that has a specified probability of containing the parameter being estimated. Most commonly used are the 95% and 99% confidence intervals that have .95 and .99 probabilities respectively of containing the parameter. The width of the confidence interval gives some indication about how uncertain we are about the unknown population parameter. Confidence intervals are more informative than the simple results of hypothesis tests (where we decide 'reject the null hypothesis' or 'don't reject the null hypothesis') because they provide a range of plausible values for the unknown parameter."

What is a Confidence Interval?

(<http://www.health.state.ny.us/nysdoh/chronic/confint.htm>)

"A confidence interval is a range around a measurement that conveys how precise the measurement is. For most chronic disease and injury programs, the measurement in question is a proportion or a rate (the percent of New Yorkers who exercise regularly or the lung cancer incidence rate). Confidence intervals are often seen on the news when the results of polls are released. This is an example from the Associated Press in October 1996:

The latest ABC News-Washington Post poll showed 56 percent favored Clinton while 39 percent would vote for Dole. The ABC News-Washington Post telephone poll of 1,014 adults was conducted March 8-10 and had a *margin of error of plus or minus 3.5 percentage points.* (Emphasis added).

Although it is not stated, the margin of error presented here was probably the 95 percent confidence interval. In the simplest terms, this means that there is a 95 percent chance that between 35.5 percent and 42.5 percent of voters would vote for Bob Dole (39 percent plus or minus 3.5 percent). Conversely, there is a 5 percent chance that fewer than 35.5 percent of voters or more than 42.5 percent of voters would vote for Bob Dole.

The precise statistical definition of the 95 percent confidence interval is that if the telephone poll were conducted 100 times, 95 times the percent of respondents favoring Bob Dole would be within the calculated confidence intervals and five times the percent favoring Dole would be either higher or lower than the range of the confidence intervals.

Instead of 95 percent confidence intervals, you can also have confidence intervals based on different levels of significance, such as 90 percent or 99 percent. Level of significance

is a statistical term for how willing you are to be wrong. With a 95 percent confidence interval, you have a 5 percent chance of being wrong. With a 90 percent confidence interval, you have a 10 percent chance of being wrong. A 99 percent confidence interval would be wider than a 95 percent confidence interval (for example, plus or minus 4.5 percent instead of 3.5 percent). A 90 percent confidence interval would be narrower (plus or minus 2.5 percent, for example)."

What does a confidence interval tell you?

The confidence interval tells you more than just the possible range around the estimate. It also tells you about how stable the estimate is. A stable estimate is one that would be close to the same value if the survey were repeated. An unstable estimate is one that would vary from one sample to another. Wider confidence intervals in relation to the estimate itself indicate instability. For example, if 5 percent of voters are undecided, but the margin of error of your survey is plus or minus 3.5 percent, then the estimate is relatively unstable. In one sample of voters, you might have 2 percent say they are undecided, and in the next sample, 8 percent are undecided. Although the difference in percentage of undecided voters may seem large, both samples have values within the margin of error of the initial survey sample.

On the other hand, narrow confidence intervals in relation to the point estimate tell you that the estimated value is relatively stable; in other words, repeated polls would give approximately the same results.

UCLA CENTER FOR HEALTH POLICY RESEARCH

Technical Assistance Series

Article #4:

What to Do When the Data You Need Is Not Available

December 2004

What Can You Do If the Data You Need Is 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. Make estimates using existing data
- C. Paint a picture
- D. Ask a researcher

A. Use Proxy Measures

Proxy measures are information that can substitute for the data you need because it relates closely to your issue. For example, if you need up-to-date poverty data for people in your neighborhood, you could take the number of people receiving Medi-Cal —since Medi-Cal is limited to low-income persons— as an indicator or proxy of the poverty level of people in the neighborhood. Note that this would be an undercount since not all low-income people are enrolled in Medi-Cal. This data does not give you a precise number or rate of your problem, but it provides useful comparisons between communities; for example, “The rate of poverty in our neighborhood may be much higher than the city average, as shown by our higher rate of people receiving 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.

B. Make Estimates Using Existing Data

Extrapolating involves taking a national, state, or county pattern of a problem and applying that pattern to your local area. Diabetes, for example, is a condition that needs on-going medical care. We know that 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 (or state or county data if you have access to it), you can take the following steps:

1. Identify the diabetes rate (proportion of people with diabetes) using the national data source. Obtain the rate for subgroups where there is variation (e.g. race, sex, age, or income). For example, the diabetes rate for Latinos nationally is .02 for ages 18-44, .143 for ages 45-64, and .203 for ages 65 and over.

[Source: National Health Interview Survey, Table 8, available 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
 - c. 2,000 Latinos ages 65 and over
3. Multiply the national rates by the local numbers and add them up.

Number of Latinos with diabetes in your community:

	RATE	x	POPULATION in your community	
Ages 18-44	.02	x	30,000 people	= 600
Ages 45-65	.143	x	11,000 people	= 1,573
Ages 65 and over	.203	x	2,000 people	= 406

Next add up the various populations with diabetes:

$$600 + 1,573 + 406 = 2,579$$

There are an estimated 2,600 Latinos with diabetes in your community

Remember that this method does not provide precise data on your topic, but it offers a way to generate estimates that can be useful in your program planning and policy advocacy work.

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. With this approach, you piece together data from several sources to illustrate your argument.

First, think about whether other communities or other issues are similar to your community and your health issue. If the available data does not apply to your intended population, health issue, or region, perhaps you could show that other communities or issues are similar. Data from other communities and issues can help you describe the demographics and issues in your community.

Next, try rethinking your search for data. How else can you approach the problem or issue you want to address? What can you do to support your argument with the data you do have?

Finally, choose a few pieces of data that are most relevant, and organize the data you have to make a convincing argument.

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 to get a response, since many researchers have moved on to a new project by the time results from a particular study become available to the public. When you do get data this way, pay attention to any limitations the researcher mentions about the data. These limitations may be the reason the researcher did not publish that information, even if he or she found it interesting. Seek out those sources of information that provide ongoing support or technical assistance.

Appendix B: AskCHIS Worksheets and Examples

WORKSHEET 1:

DEFINING THE DATA QUERY & POPULATION AND PREPARING THE DATA QUERY

Step 1: What question are you trying to answer? Or, what topic(s) are you interested in studying?

For example: *What percentage of adults in California has health insurance?*
Write your own research question(s) below.

Step 2: Review the CHIS fact sheet, explore the list of topic categories, or use the keyword search tool (in the **Select a MAIN TOPIC** page in AskCHIS) to check the availability of health topics.

Step 3: Identify the health topic(s) you are interested in studying.

- For each topic selected, use the **Keyword Search** tool (or the **View Topics by Category** feature) to find the variables related to the research question you identified in Step 1.
- For example, someone interested in health insurance issues may chose the variable "Currently insured."
- Write out the best variable for your query below.

Health insurance

a) Topic of interest: _____

Currently insured

Best CHIS Variable: _____

b) Topic of interest: _____

Best CHIS Variable: _____

Step 4: Choose the population(s) included in your results. In other words, what population will your data query describe?

Select one or more options from each of the demographic items. This information will be used to create and submit your data query.

a) Age:

Children (define age groups) _____ to _____ years

Adolescents (define age groups) _____ to _____ years

Adults (define age groups) _____ to _____ years

Entire population (define age groups) _____ to _____ years

b) Race/Ethnicity:

Hispanic / Latino _____

American Indian/Alaska Native _____

Asian _____

African American _____

White _____

Other single/2 or more races (i.e. 2 or more races) _____

All racial groups _____

c) Gender:

Male _____

Female _____

d) Federal Poverty Level (FPL):

0-99% of FPL _____

100%-199% of FPL _____

200%-299% of FPL _____

300% or more of FPL _____

All income levels _____

e) Geographic region (required)

The Entire State (i.e. all counties in California) _____

Or identify county or counties of interest _____

WORKSHEET 2:

Query for a 2 Variable Table (Bivariate)

Step 1: What question are you trying to answer? Or, what topic are you interested in studying?

For example: In California, what percentage of Latinos have been diagnosed with asthma, and what percentage of African Americans have been diagnosed with asthma?

Write your own research question(s) below.

Step 2: Review the CHIS fact sheet, explore the list of topic categories, or use the keyword search tool (in the **Select a MAIN TOPIC** page in AskCHIS) to check the availability of health topics.

Step 3: Identify the health topic(s) you are interested in studying.

- For each topic selected, use the **Keyword Search** tool (or the **View Topics by Category** feature) to find the variables related to the research question you identified in Step 1.
- Write out the best variable for your query (e.g., *Ever diagnosed with asthma*).

a) Topic of interest: _____

Best CHIS variable: _____

b) Second (comparison) variable:

Write down a variable you want to use for comparing your results. Common choices include race, gender, Federal Poverty Level, and current health insurance. Use the **Keyword Search** tool or the **View Topics by Category** feature to find a comparison variable (e.g., *Race-UCLA CHPR*).

Best CHIS comparison variable: _____

Step 4: What population are you interested in learning about? Or, what population will your data query describe?

Select one or more options from each of the demographic items. This information will be used to create and submit your data query.

a) Age:

Children (define age groups) _____ to _____ years
 Adolescents (define age groups) _____ to _____ years
 Adults (define age groups) _____ to _____ years
 Entire population (define age groups) _____ to _____ years

b) Race/Ethnicity:

Hispanic / Latino _____
 American Indian/Alaska Native _____
 Asian _____
 African American _____
 White _____
 Other single/2 or more races (i.e. 2 or more races) _____
 All racial groups _____

c) Gender:

Male _____
 Female _____

d) Federal Poverty Level (FPL):

0-99% of FPL _____
 100%-199% of FPL _____
 200%-299% of FPL _____
 300% or more of FPL _____
 All income levels _____

e) Geographic region (required)

The Entire State (i.e. all counties in California) _____
 Or identify county or counties of interest _____

Step 5: Go to the AskCHIS website.

WORKSHEET 3: STRATEGIES FOR HANDLING UNSTABLE ESTIMATES & WORKING WITH DATA FOR SMALL COUNTIES

Unstable Estimates in AskCHIS:

- AskCHIS provides the option to view data at the county level and at the region (multi-county) level.
- In many cases uses county level data limits the amount of comparison or “drill down” a user can do without generating query results that include unstable estimates.
- Given the small size of some California counties, data pulled for certain health indicators can yield “statistically unstable” results.
- In AskCHIS, statistically unstable results are marked by a red asterisk (*) as seen below.

	Race - OMB/Department of Finance	
Ever diagnosed with asthma	Latino <i>Sort by this</i>	<i>Sort by this</i>
Has asthma	9.8%* (0.0 - 24.4) 2,000	60.3%* (7.5 - 100.0) 6,000
Does not have asthma	90.2%* (75.6 - 100.0) 18,000	39.7%* (0.0 - 92.5) 4,000

Unstable Estimates are Usually Caused By:

- Small sample size available for certain categories:
 - Example: Not enough Latino adults with diabetes were sampled in order to report a rate
- The coefficient of variance (CV) was too large:
 - Example: a small sample of individuals in a particular CHIS table cell can lead to a great deal of variance between responses. CHIS data is not reported with any CV larger than .3.

Step 1: Execute a query by selecting a specific geographical area, main topic, compare by group and target population. List the criteria used:

Geographical Area: _____

Main Topic: _____

Compare By: _____

Population: _____

Step 2: The query should result in a one by one or two by two table. Do any of the cells in your table have a red asterisk next to the CHIS estimate given? _____

If so, where in your table was the unstable estimate warning? _____

Note what selections from Step 1 may be contributing to the small sample size.

Any unstable estimates (red asterisk marked values) in your table, are areas where the sample size is low and the variability is high. There are several strategies for making your query more general, thus increasing the number of survey participants in each category.

Strategies for Stabilizing Unstable Estimates in AskCHIS:

- Increase/change the geographic area or use statewide totals
- Remove comparative categories (such as race/ethnicity, gender, age)
- Use data from a larger time frame. Combine information from a series of years instead of only using one survey.
- If the rates of disease or injury in a community are unstable, it is sometimes helpful to use other indicators to look at overall health status of the population.

Step 3: Try each of the following suggestions to see the impact of changing your query criteria on the stability of the estimates generated:

- 1) Combine years of CHIS data used to make for a larger sample size by selecting the "Change Time Period" option on the results page.
Estimate now stable? (Y/N) _____
- 2) Increase the geographical area used in the query: Include multiple counties or regions, or even default to statewide data.
Estimate now stable? (Y/N) _____
- 3) Remove any gender or age categorization from the Compare By section.
Estimate now stable? (Y/N) _____
- 4) Choose a different related indicator, such as different measures for management of the same disease.
Estimate now stable? (Y/N) _____

Step 4: If estimate is still unstable, continue to use more general populations and geographic areas until estimate is stable. Report and cite only statistically stable estimates from AskCHIS. If you have difficulty generating stable estimates, please contact Health DATA's technical assistance team for further instruction.

WORKSHEET 4: INTERPRETING AND APPLYING THE DATA RESULTS

Step 1: Who will use the data obtained from your query?

Possible Audiences

- Media_____
- Program Planners_____
- Advocates_____
- Policy Makers_____
- Health Care Providers_____
- Grant Writers/ Fundraising/Development_____

Others? (List below)

Step 2: What do you hope to accomplish with the data?

Example: Our organization seeks to obtain funds to provide a health promotion program in the community.

Step 3: Below, write your summary of the results.

For a One-Way Table:

Fill in the blanks below (not all demographic items are required):

In _____, _____% of _____, _____, _____, _____ have/are _____.

Region Percent Poverty Race/Ethnic, Gender, Ages, Main Topic.
 from AskCHIS Level Group

APPENDIX C: DATA RESOURCES

Data Resources

Below are some resources - in alphabetical order by topic - which you can refer to for additional information to help you search for, understand, and apply relevant data. Websites are subject to change without notice or may require registration.

ASSET MAPPING:

1. University of Wisconsin-Madison. Identifying, Mapping, and Mobilizing Our Assets. <http://www.uwex.edu/ces/lmcourse/PDFs/assets.pdf>.
2. Northwest Regional Educational Laboratory. Mapping Community Assets Workbook. http://www.nwrel.org/ruraled/publications/com_mapping.pdf
3. Healthy-Children-Healthy City Asset Mapping Project. <http://www.healthycity.org>

COMPUTER RESOURCES:

1. Center for Nonprofit Management - <http://www.cnmsocal.org/>
2. Zero Divide/Community Technology Foundation of California: Social justice, access & equity through computer technology - <http://zerodivide.org/>
3. Techsoup: The Technology Place for Nonprofits - <http://techsoup.org/>

DATA REPORTS AVAILABLE TO THE PUBLIC:

1. Berkeley-Charleston-Dorchester Council of Governments and Trident United Way. Countywide Needs Assessment for the Trident Area of South Carolina. www.tuw.org/needs/titlepage.html
2. Healthcare Association of Southern California and the County of Orange Health Care Agency. Orange County Health Needs Assessment. www.ochna.org

GRANT WRITING / SEEKING FUNDING:

1. The Foundation Center. Helping grant seekers succeed, helping grant makers make a difference. <http://fdncenter.org/>
2. The Foundation Center. Proposal Writing Short Course. <http://fdncenter.org/learn/shortcourse/prop1.html>
3. Michelle K. Carter. Organizing the Grant Writing Process: 7 Things to Do Before Writing (in English and Spanish). <http://www.scn.org/ip/cds/cmp/modules/res-grn.htm>
4. North Central Regional Educational Laboratory. Findings and Developing Resources for Comprehensive School-Linked Strategies for Children and Families. <http://www.ncrel.org/sdrs/areas/issues/envrnmnt/css/ppt/chap3.htm>
5. Partners in Information Access for the Public Health Workforce. Links to Grants and Funding Opportunities. <http://phpartners.org/grants.html>
6. Phil Bartle. Community Project Resources (in English and Spanish). <http://www.scn.org/ip/cds/cmp/modules/res-int.htm>

ONLINE DATA QUERY SYSTEMS:

1. Centers for Disease Control and Prevention. **Behavioral Risk Factor Surveillance System**. <http://www.cdc.gov/brfss>

BRFSS provides data for all 50 states on behaviors related to the leading causes of morbidity and mortality. The online query system provides access to state- and national-level data from 1995-2003. Limited local area data is also available via SMART BRFSS (Selected Metropolitan/Micropolitan Area Risk Trends) project.

2. UCLA Center for Health Policy Research. AskCHIS: The Online Query System for the California Health Interview Survey. <http://www.askchis.com>

AskCHIS is tool to search for data from the California Health Interview Survey, the largest state health survey in the country. Information on particular health conditions, health-related behaviors, access and use of health

services, health insurance, and participation in public programs is included. Data is available at the state, region, county, and Los Angeles Service Planning Area level.

RAW DATA:

For **ADVANCED** data users who are comfortable doing their own data analysis using computer programs such as SAS, SPSS, Microsoft Excel, or Microsoft Access:

The U.S. Census Bureau and the Centers for Disease Control and Prevention. **The Data Web/DataFerrett**. <http://www.thedataweb.org/>

DataFerrett is a free, downloadable program that allows you to access 15 datasets from large, national surveys. Datasets include the American Community Survey, Current Population Survey, Decennial Census, National Ambulatory Medical Care Survey, and the National Health and Nutrition Examination Survey.

REPORT WRITING:

Phil Bartle. **Report Writing**. <http://www.scn.org/ip/cds/cmp/modules/rep-int.htm>

USING DATA IN PLANNING AND PRACTICE:

1. Phil Bartle. **Preparing a Community Action Plan: The Community Decides its Own Future**. <http://www.scn.org/ip/cds/cmp/modules/org-cap.htm>
2. The Access Project. **Using Data: A Guide for Community Health Activists**. <http://www.accessproject.org/downloads/data.pdf>