

### Strengths and Limitations of AskCHIS, AskCHIS NE and CHIS

#### **Strengths/Benefits and Limitations of CHIS Data:**

It is important to evaluate any secondary data source for limitations before you decide to use it. In this case, you will need to decide when to use CHIS data and specifically when to use the AskCHIS© or AskCHIS© Neighborhood Edition (NE) tools. You can do this by thinking about secondary data in terms of the following five evaluation criteria. Since we will be discussing CHIS data, the examples below apply these evaluation criteria to CHIS.

#### **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 I 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?”

Data collection sources have varying levels of credibility based on their interest to present 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 center data are similar to government statistics, in part because they undergo peer review.

**Example:** CHIS data maintain a high level of credibility because data are collected and produced by the UCLA Center for Health Policy Research, a unit within the UCLA School of Public Health. It 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 public 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 the knowledge of health issues.

#### **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 ask yourself, “How do the goals or content of the study relate to my needs? Are the conditions of the study unique to a particular case? How close is the relationship between my needs 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 data item to ensure its specificity to the issue you want to address. 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 and country of birth.
- Geography: CHIS data are collected from all 58 counties in California across 44 geographic areas that represent 41 individual counties and three groupings of counties with smaller populations.
- Other CHIS Topics (varies across CHIS cycles): health status, health conditions, mental health, health behaviors, women's health, dental health, food insecurity/hunger, food environment, neighborhood

and housing, access to use of health care, health insurance, public program eligibility, parental involvement, child care, employment and income.

### **Generalizability – Can you apply data from one population to describe another population?**

Generalizability refers to specific data and how well it can be used to describe other populations. You may ask yourself, “What are the characteristics of the secondary data’s population/sample? Do the participants provide data that can be applied to 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 can be generalized at the state and county levels. Rigorous sampling techniques ensure sufficient amounts of data are collected to describe populations throughout the state. For instance, starting in 2001, CHIS data is collected on a continuous survey cycle year via random-dial telephone surveys with over 40,000 to 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.

### **Reliability - How was the data collected?**

Reliability refers to the accuracy of the data. You may ask yourself, “Can I trust the data 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 highly reliable because the highest research practices and standards are applied. Below are a few methods used to ensure data quality:

- 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 landline and cellular telephone numbers from each geographic area sampled with a minimum number of people included.
- To ensure that public health trends are accurately recorded over time, CHIS selects and surveys unique households during each survey cycle.
- CHIS randomly selects and interviews only one adult per household. Only that selected person can participate.
- If there are children in the household, CHIS also asks questions about an adolescent (ages 12–17) and a child (ages 0–11).
- To ensure the inclusion of California's diverse racial and ethnic populations, thousands of CHIS interviews are conducted in languages other than English, including Chinese (Cantonese and Mandarin), Korean, Spanish, Tagalog and Vietnamese.

### **Timeliness – When was data collected?**

Timeliness refers to when the research was conducted. You may 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 collection cycles occur frequently and are conducted every year with limited annual estimates available. The high frequency of collection and production of CHIS data increases the relevancy and timeliness of data findings and trends.

If you are unsure about the strengths, benefits and limitations of a secondary data source, go to the source and ask them more details, if possible.

No data is perfect. 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 is very useful for making powerful statements about community health. However, CHIS data does 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 data *CANNOT* 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 2012).
- How does the number of women with breast cancer compare to women from other states and nationally? (CHIS only collects data on California households).

**AskCHIS NE© Data Limitations and Use:**

Most health estimates available in AskCHIS© NE are model-based on small area estimates (SAEs), such as a zip code. They are created by statistical models of relationships using CHIS data at a larger geographic level and then applying those models to local population data. These complex statistical models use relevant characteristics of populations and geographic areas to predict health conditions for small geographic units (cities, zip codes, for example).

SAEs are not direct estimates (estimates produced directly from survey data, such as those provided through AskCHIS©). While direct estimates are produced solely using survey data and design weights, the model-based estimates in AskCHIS© NE also rely on secondary data describing characteristics of both geographic regions and populations.

AskCHIS© NE users should decide on the appropriateness of using model-based SAEs based on the strengths and limitations briefly discussed below. For more information, methodology documentation is made available through the AskCHIS© NE website.

In developing final local-level estimates, data from CHIS were used as the primary data source for modeling. Area-level data providing contextual information was provided by the Census Bureau's American Community Survey (ACS). Population characteristics were provided by Nielsen-Claritas Pop-Facts (Claritas) data.

Like all estimates, the health estimates in AskCHIS© NE are subject to errors that can impact data accuracy. These include sampling error and non-sampling error of the input data, and model error.

### **Sampling Error**

Sampling errors occur because inferences about the entire population are based on information obtained from only a sample of that population. If the sample is representative of the target population, as with CHIS, then sampling error is reduced. The models for AskCHIS© NE health estimates are built on data samples of California residents from CHIS and ACS instead of information from all members of the California population.

### **Non-Sampling Error**

Non-sampling errors include coverage errors, measurement errors (respondent, interviewer, questionnaire or collection method), non-response errors and processing errors. CHIS and ACS data, as with all survey data, are subject to these errors. Non-sampling errors are partially corrected through post-collection data cleaning and weighting processes. Claritas data are based on Census data and administrative data, and thus also subject to non-sampling errors.

### **Model Error**

Generally, model errors occur when the statistical model does not account for all the information contributing to variation of the dependent variable. In AskCHIS© NE, model error is reduced by borrowing strength from several data sources to inform the statistical models.

### **Notice Regarding Demographic Variables**

Socio-demographic variables available in AskCHIS© NE were produced using data from the American Community Survey (ACS). These data were adjusted to match the CHIS population, which excludes populations living in group quarters (such as prisons, hospitals or dormitories). The demographic variables are included in AskCHIS© NE to provide additional context to health estimates and may not be generalized to the entire population of California.

AskCHIS© NE is a public health surveillance tool, not an official source of demographic information. Demographic information available in AskCHIS© NE is not meant to replace data from the U.S. Census.

### **Liability Disclaimer:**

Your use of estimates, data and features from AskCHIS© NE signifies agreement that the Regents of the University of California, UCLA, the UCLA Center for Health Policy Research and The California Health Interview Survey shall not be liable for any activity involving these data, estimates or features of them for any purpose.