CHIS Researcher Workshop

UCLA Center for Health Policy Research &
UCLA Statistical Consulting Group

February 28, 2014
Presentation Agenda

1. CHIS Overview
   a. History and purpose
   b. Sample size and characteristics
   c. Survey methods
   d. Survey content

2. Access to CHIS data
   a. Public Use Files (PUF)
   b. AskCHIS
   c. Data Access Center
Presentation Agenda

3. Using CHIS Data
   a. Data analysis examples
      i. Weighting and variance estimation
      ii. Replicate weights and Taylor series
   b. Pooling CHIS years
   c. Data linkage and GIS
   d. Imputation and multiple imputation
Part 1:
Overview: What is CHIS?

David Grant
CHIS Director
CHIS history

- Conceived out of need for better data on health insurance coverage in California, Statewide and counties
- Three-year planning process began in 1996/7
- First fielded in 2001 and conducted as a biennial survey through 2009 and continuously beginning in 2011-2012
Purpose & mission

- CHIS is California’s source of state and local population-based health data

- CHIS is designed to provide easy access to data and estimates:
  
  1. To support decision making at the local level and statewide in public health and health care
     - For policy analysis, development and advocacy
     - For State and county surveillance of public health indicators
     - For service and program planning, development, and evaluation
  
  2. To understand and measure health needs and disparities in California — characterized by ethnic, geographic, and social class diversity
Purpose & mission

- Funded by federal and state health agencies, California and national foundations, and others
- CHIS data is widely used by
  - State agencies
  - County health departments
  - Academic researchers and students
  - Advocacy groups and CBOs
- Covers non-group quarters population in California
  - Adults (age 18 and over)
  - Adolescents (ages 12 to 17)
  - Children (ages 0 to 11 by parental proxy)
Methods: how is CHIS done?

- Large, geographically stratified random digit dial (RDD) telephone survey designed to provide statistically reliable estimates:
  1. At the local level for counties (adults) and statewide
  2. For major racial/ethnic groups and many ethnic subgroups
- Geographic stratification ("complex" sample)
  - 56 total geographically defined strata
    - 2 counties with sub-county strata
      - Los Angeles: 8 Service Planning Areas (SPAs)
      - San Diego: 6 Health Regions
    - 39 individual county strata (excluding LA and San Diego)
    - 3 grouped county strata containing the 17 smallest counties (by population)
Methods: how is CHIS done?

- Two separate RDD samples:
  1. Landline telephone numbers at stratum level (56 strata)
  2. Cellular telephone numbers (since CHIS 2007) at “region” level (7 regions)
- Ethnic oversamples
  - Koreans and Vietnamese oversampled in each CHIS cycle through:
    - Geographic targeting
    - Surname list sample
- Interviews conducted in 7 languages: English, Spanish, Chinese (Cantonese and Mandarin), Korean, Vietnamese with Tagalog added in 2014
CHIS sample

- See CHIS Sample Size by Age and Race/Ethnicity – see handouts

<table>
<thead>
<tr>
<th>Age Group</th>
<th>2001</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
<th>2011-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult (18+)</td>
<td>56,270</td>
<td>42,044</td>
<td>43,020</td>
<td>51,048</td>
<td>47,614</td>
<td>42,935</td>
</tr>
<tr>
<td>Child (0-11)</td>
<td>12,802</td>
<td>8,526</td>
<td>11,358</td>
<td>9,913</td>
<td>8,945</td>
<td>7,334</td>
</tr>
<tr>
<td>Teen (12-17)</td>
<td>5,733</td>
<td>4,010</td>
<td>4,029</td>
<td>3,638</td>
<td>3,379</td>
<td>2,799</td>
</tr>
</tbody>
</table>
What information does CHIS collect?

- See CHIS Questionnaire Topics – see handouts

- Rich demographic data
  - Age, sex, race/ethnicity, marital status
  - Country of birth, years in US, citizenship, mother/father’s country of birth, language/s spoken at home, English proficiency
  - Employment status, income, poverty level, educational attainment

- Health behaviors
  - Alcohol consumption and binge drinking
  - Tobacco use (past and current)
  - Illegal drug use (teens)
  - Physical activity, dietary intake
  - Cancer screening, flu shot, HIV/STD testing, use of seat belts and bike helmets
What information does CHIS collect?

- Health conditions:
  - General health
  - Chronic conditions such as asthma, diabetes, heart disease, high blood pressure, COPD, etc.
  - Mental health
  - Disability

- Access to and use of health care services
  - Usual source of care
  - Visits to doctor, ER use
  - Communication with doctor, need for translator, health literacy
What information does CHIS collect?

- Health insurance coverage:
  - Current coverage
  - Coverage type (employer-based, private purchase, MediCal, Healthy Families, MediCare)
  - HMO enrollment and plan details
  - Coverage past 12 months

- Uninsured
  - Reasons for lack of coverage
  - Medi-Cal and Healthy Families eligibility (age < 65)
What information does CHIS collect?

- Other topics & information
  - Food security and hunger
  - Public program participation
  - Geocoding of households and schools
  - Years lived at address and in neighborhood
  - Social relationships and neighborhood conditions (2003)
  - Family history of cancer (2005)
  - Inter-personal violence (2007 & 2009)

- ... and numerous other topics
Part 2: Access to CHIS Data

Sansan Lin
Health Research Privacy Compliance Coordinator
Channels of Data Access

1. Public Use File (PUF)
2. AskCHIS
3. Data Access Center (DAC)
Public Use Files

- Available at no cost for download through the CHIS web site (chis.ucla.edu)
- Statewide files (no sub-state geographic identifiers)
- Excludes highly sensitive variables like mental health and citizenship status
- Separate files for adults, adolescents, and children
- Available in SAS, SPSS, and STATA format
Public Use Data Files

Conduct your own in-depth research

Dig deep into California health issues with our comprehensive statewide CHIS data files on a variety of topics. Public Use Files (PUFs) enable researchers to customize and run their own data searches. The files are available in a variety of data formats, including SAS, SPSS, and STATA data formats.

Users must login to view the data files.

The files contain:
- Records of individual respondents.
- Various health status, health conditions, health-related behaviors, health care access and health insurance coverage information.
- Gender, age, race/ethnicity, urban/rural and other descriptors.
- A data dictionary and survey questionnaire accompany each file.

Note: To minimize the risk of indirect identification and increase data confidentiality, sub-state geographic identifiers (e.g., county, city, and zip code) and confidential variables such as sexual behavior are excluded from the CHIS Public Use Files. However, access to confidential data is available through our Data Access Center. Learn more about confidential data.

For research requiring county, region, or state analyses visit AskCHIS.
Public Use Files

- Data files – adult, adolescent, and children—for each CHIS cycle
- Data dictionary for each file
  - Summary of survey methodology
  - Indexes for variables
  - Information about each variable including name, description, universe
- Questionnaires
AskCHIS

- ask.chis.ucla.edu
- Free use interactive data query system, available 24-7 via the Internet
- Custom and instant queries of the CHIS data
- Allows for analysis at the state, region, and county level of various health variables and demographic characteristics
- Includes sensitive or identifiable variables, unlike PUF
AskCHIS disclosure limitations

- Use of only up to two variables at a time
- Excludes certain variables pertaining to illegal activities or variables containing detailed categories that could potentially harm or embarrass a respondent
- Weighted population estimates and percentages with confidence intervals
- Rounds all estimates to the nearest 1,000
- Suppresses cells sizes of less than 500 in tables
Create an AskCHIS account. Note that your AskCHIS account and PUF account are different and you must have two different passwords.
1. Select a geographic area

- Entire State of California
- Specific large regions (i.e., groups of counties) in California
- Specific counties in California
- Los Angeles County Service Plan Areas (SPA)
- San Diego County Health Regions

NOTE: Because of the statistical analysis involved with making estimates, there may be some variance between the sum of the SPAs and Los Angeles County estimates as a whole.

View map of Service Plan Areas (SPA in Los Angeles County)
2. Select a main topic

3. Select a compare by topic (optional)
4. Select a population
5. Get instant results
AskCHIS training

- AskCHIS online tutorial and FAQ
- Online workshops

http://healthpolicy.ucla.edu/programs/health-data/trainings/Pages/askchis.aspx
Data Access Center (DAC)

- DAC – a secure physical space and a secure data network
  - Maintains all CHIS data
  - Designed to protect the confidentiality of CHIS respondents

- How to Access Confidential Data
  - Prospective researchers submit a DAC application, which is reviewed by the Data Disclosure Review Committee
  - Criteria for project approval
    - Feasibility
    - Risk of disclosure of confidential information
    - Variable selection
  - Cost estimates vary from project to project and depend on the services needed
Data Access Center (DAC)

- Analyses conducted through the DAC via two ways:
  1. By utilizing the statistical and programming services of the Center’s Statistical Support and Programming staff, who will write and run codes and return results to the researcher
  2. By writing programming code for statistical analysis and submitting it to the DAC, where statisticians will run the codes and return results to the researcher
Data Access Center (DAC)

- For more information, visit the CHIS website pages on Confidential Data (http://healthpolicy.ucla.edu/chis/data/Pages/confidential.aspx)
- Contact the DAC Coordinator at dacchpr@ucla.edu for an application or to discuss questions or concerns