

## THE CALIFORNIA HEALTH INTERVIEW SURVEY (CHIS)

# The California Health Interview Survey: Balancing Respondent Confidentiality with Access to Data

David Grant, PhD

UCLA Center for Health Policy Research  
Director, California Health Interview Survey

# This presentation

- **Very brief overview of the California Health Interview Survey**
- **Types of data and information collected**
- **Strategies for disseminating data while protecting respondent confidentiality**
  - **Publications**
  - **Public use files (PUFs)**
  - ***AskCHIS* Internet query system**
  - **Data Access Center**

# CHIS introduction

- **The California Health Interview Survey (CHIS) is California's source of state and local population-based health data**
- **Conducted every other year since 2001 and continuously since 2011**
- **Large sample of households interviewed every 2 year data collection cycle: 42,000 to 56,000 households per cycle (about 1 in every 260 households)**
- **Conducted by the UCLA Center for Health Policy Research**
- **Sponsored by State and Federal agencies and foundations**

# CHIS introduction

- Interviews are conducted by telephone through a random digit dial sampling method of landline and cellular telephone numbers
- Sample is geographically stratified to produce independent estimates for most of California's 58 counties
- Separate interviews are conducted with adults (ages 18+), adolescents (12 to 17) and about children (ages 0 to 11)
- Interviews are conducted in English, Spanish, Chinese (Cantonese & Mandarin), Korean, and Vietnamese

# What information does CHIS collect?

- **Health behaviors**
- **Health conditions**
- **Access to and Utilization of health care services**
- **Health insurance coverage**
- **Uninsured**
- **Mental health and access/use of mental health services**
- **And many other topics...**

# What information does CHIS collect?

- **Rich demographic data, including:**
  - Age, sex, race/ethnicity, ethnic origin/ancestry, marital status, sexual orientation
  - Country of birth, years in US, mother/father's country of birth
  - Language of interview, language/s spoken at home, English proficiency
  - Employment status, income, poverty level, educational attainment, Veteran's status
  - Citizenship status (citizen, green card, pending, other)
  - Respondent's geocode (latitude and longitude)
  - Ability to link adult, teen, and child records

# What info does CHIS disseminate?

- **With caution—as much as we can!**
  - 20% of total CHIS budget dedicated to dissemination
  - Attempt to “democratize access to data”
- **CHIS data is self-report and is not covered by HIPAA; Oversight of CHIS dissemination is governed by UCLA and State of Calif. HHS Institutional Review Boards**
- **Strategy to disseminate information and data in multiple formats to reach different audiences**

# CHIS dissemination

- **AskCHIS**—Internet query system; very popular with 40,000 registered users making 800,000 queries of CHIS data
- **Public Use Files (PUF)** – available for free download from the CHIS website
  - Adult, teen, and child PUFs in SAS, SPSS, and STATA format
- **Data Access Center (DAC)** – provides access to confidential data for advanced statistical users with complex data needs and analytic capabilities
  - Sensitive & identifiable variables only available through the DAC, including access to geocoded data



# Protecting respondent confidentiality

- **Overarching strategies:**
- **Obtain a Certificate of Confidentiality from the National Institutes—protects against subpoena of respondent information**
- **Protocols in place so that the most identifiable information (telephone number, names, addresses) is never received by UCLA from data our collection subcontractor**
- **All confidential data resides in our Data Access Center, a segregated, HIPAA compliant network (State audited and approved)**

# Protecting respondent confidentiality

- **Public Use Files (PUFs)—available for free download from web**
  - Exclude all geographic identifiers
  - Exclude potentially identifiable information (exact date of birth)
  - Exclude sensitive content (e.g., citizenship, sexual identity, mental health)
  - Cannot link adult, child, and teen data
  - Top and bottom coding of continuous variables (e.g., age, height)
  - Data perturbation of ‘population uniques’
  - User acknowledgement of terms and use of data before downloading

# Protecting respondent confidentiality

- **AskCHIS – provides access to county level data**
  - All estimates are weighted to the county/state population and rounded to the nearest 1,000
  - All variables are categorical
  - Algorithms suppress small sample sizes
  
- **Data Access Center (DAC) – access to confidential data**
  - Sensitive & identifiable variables only available thru the DAC
  - Application process and Data Disclosure Review Committee
  - Protocols govern release of confidential data and output



# THE CALIFORNIA HEALTH INTERVIEW SURVEY (CHIS)

**Thank you**