

UCLA Center for Health Policy Research Data Access Center

M. Leeann Habte
Data Access Manager
UCLA Center for Health Policy Research

Committee on Review of National Immunization Program's Research Procedures
and Data Sharing Program
Meeting One
August 23, 2004



www.healthpolicy.ucla.edu

Overview



- Background on Data Access Center
- Process for accessing California Health Interview Survey (CHIS) data
 - Role of CHIS Data Disclosure Review Committee
- Data Access Center operating procedures
 - Disclosure review process

Background



- Data Access Center
 - Secure, controlled environment where researchers can access confidential data under supervision of DAC staff
 - Modeled after the National Center for Health Statistics' Research Data Center

- Current confidential data source
 - California Health Interview Survey (CHIS)
 - California's assessment tool designed to meet state and local needs for population-based health data
 - Very large biennial survey of California population
 - Estimates for many health and social indicators and access to health care

Background



- CHIS data are used:
 - To support decision making at local level and statewide in public health and health care
 - For policy analysis, development and advocacy
 - For service and program planning and fund development
 - To understand and measure health needs of California's population — characterized by ethnic, geographic, and social class diversity
 - For policy analysis, development and advocacy
 - For research

Background



- Facility security
 - Electronic security lock for room
 - Multiple workstations supported by data server on private network with secure printer
 - No Internet access
 - Floppy disc and CD Rom drives are disabled
 - Printer and server located in separate room accessible to senior staff only

Background



- Restrictions on access to data
 - Access only for researchers with approved projects
 - Access only to the data needed to complete their projects
- Use of data supervised by DAC staff
 - DAC staff member on site when guest researcher is present
 - All materials inspected by DAC senior staff prior to removal
- Output restricted
 - Disclosure review conducted on all output by senior staff
 - No record level information or micro data released
 - No tables or frequencies with small cell sizes released

Background



- Ways to access data
 - Work on site as guest researcher
 - Use DAC programming services
 - E-mail computer programs to DAC staff to be run with confidential data

- Analytical software available
 - SAS, Stata, SPSS
 - SUDAAN, Wesvar
 - Stat Transfer
 - Custom software installed upon request

Background



➤ History

- Began operating in October, 2003 to provide access to CHIS data
- 58 independent research projects have been approved for use of CHIS 2001 data
 - 25 led by faculty associated with UCLA Center for Health Policy Research
 - 33 led by researchers from other institutions or organizations
 - 24 by California academic or research institutions
 - 9 by universities outside California

Background



- Services used by projects external to Center
 - 8 projects used guest researcher service
 - 4 used DAC programming services
 - 21 e-mailed computer programs for DAC staff to run
- Costs
 - \$500 initial set-up fee
 - \$65/hour for guest research access
 - \$140/hour for programming services
 - \$120/hour to run programs
 - \$1,000 minimum fee per project
 - Charges based on actual time spent on project

Background



➤ CHIS Research Clearinghouse

- Provides information about research projects using CHIS data
- Designed to promote collaboration and information sharing
- More than 100 projects currently listed
- See www.chis.ucla.edu/rc for details on projects using CHIS data



Process for accessing CHIS data

- Researchers submit applications to the DAC
 - DAC staff prepares summary of application
 - Applications sent to CHIS Data Disclosure Review Committee
- Applications include:
 - DAC application forms
 - Personal and organizational information
 - Service request
 - Abstract
 - DAC Nondisclosure Affidavit requires agreement to:
 - Protect confidentiality of respondents
 - Acknowledge CHIS
 - Submit copies of publications to DAC



Process for accessing CHIS data

- Supplemental materials
 - Biographical sketch or resume
 - List of CHIS variables requested
 - Researcher must request variables using DAC variable lists
 - Detailed description of any user-supplied files
 - Copy of approval or exemption by home institution's Institutional Review Board
 - Evidence of faculty sponsorship (for students)
 - Research Clearinghouse on-line submission (optional)

Process for accessing CHIS data



- CHIS Data Disclosure Review Committee
 - Technical committee that implements confidentiality policies and procedures adopted by CHIS Governing Board
 - Reviews data files, data products, and requests for access to CHIS data
 - Includes members of the UCLA Center for Health Policy Research, California Department of Health Services, and Public Health Institute
 - Meets bi-weekly to review new applications
 - Makes recommendations to the CHIS Principal Investigator for approval, request for further information, or denial of application



Process for accessing CHIS data

➤ Review Criteria

➤ Consistency with purpose of CHIS

- Is proposed use of data consistent with the purpose for which data was collected?

➤ Feasibility

- Is sample size sufficient?
- Are CHIS data appropriate for answering the research questions proposed?
- Are variables requested related to proposed analyses?



Process for accessing CHIS data

➤ Review Criteria

➤ Risk of disclosure of confidential information

➤ Does project propose to merge user-supplied data with CHIS data?

➤ Geographic level?

➤ What additional risks of disclosure are associated with the merged data set?

DAC operating procedures



➤ Project Initiation

- A custom data set is created for the project, based on the approved list of variables
- Researcher is provided with limited technical assistance on CHIS variables, weighting, and variance calculation
- A cost estimate is developed for the requested services, and approved by the researcher
- A project number and senior programmer contact are assigned
- Dummy data files are sent to the researcher

DAC operating procedures



➤ Project Implementation

- The project is entered into the tracking data base
- Project-specific timelines are developed for programming services
- Computer programs e-mailed to the DAC are run within five working days of receipt
- Time spent on services for the project is tracked
- Researcher is billed on a quarterly basis



DAC operating procedures

➤ Disclosure review

- DAC manager or senior programmers must conduct disclosure review of all output prior to removal from DAC
 - Electronic copies or printouts

➤ First Steps

- DAC staff member summarizes primary disclosure risks associated with project
- Provides list of all CHIS variables, which are ranked numerically according to:
 - Risk of contributing to disclosure of a person's identity
 - Sensitivity level



DAC operating procedures

- Disclosure review prevents release of:
 - Tables with small cells with some level of disclosure risk associated with variables
 - Record-level information

- State-level or Los Angeles County disclosure review rules
 - Unweighted frequencies suppressed if $n < 3$
 - Weighted frequencies suppressed if $n < 5$
 - Cross-tabulation includes sensitive variable and “key” variable
 - Cross-tabulation includes “key” variable and other identifying variable



DAC operating procedures

- Stratum level disclosure review rules
 - Unweighted frequencies suppressed if $n < 3$
 - Weighted frequencies suppressed if $n < 5$
 - Cross-tabulation includes sensitive variable and identifying variable
 - Cross-tabulation includes “key” variable and other identifying variable

- Standard exceptions
 - Univariate frequencies for variable response categories
 - Small cell sizes for responses of “other,” “skipped,” or “don’t know”
 - Cross-tabulations or tables with no sensitive or identifying variables

DAC operating procedures



- Cell suppression rules
 - If there are many small cells, programmer should recommend recoding of variables
 - If there are few small cells in output, programmer must suppress small cell and do complementary suppression
- Special attention should be paid to:
 - Analyses including sensitive variables
 - Nature of sequential analyses

Acknowledgements



- CHIS Principal Investigator
 - E. Richard Brown, Ph.D., Director, UCLA Center for Health Policy Research
- Members of CHIS Data Disclosure Review Committee
 - Wei Yen, Ph.D., Associate Director, CHIS
 - Hongjian Yu, Associate Director, Statistical Support and Programming, UCLA Center for Health Policy Research
 - Lee Habte, M.A., Data Access Manager (Committee Chair), UCLA Center for Health Policy Research
 - Laura Lund, M.A., CHIS Coordinator, California Department of Health Services
 - Michael Quinn, Chief, Office of Health Information and Research, California Department of Health Services
 - Elaine Zahnd, Senior Research Scientist, Public Health Institute