

Data Management Methods

Methods for Maintaining Longitudinal Population Health Studies

*to jump to Hospital-level race/ethnic data quality DataBooks, please click [HERE](#)

1. Basic Computing Environment ^[1]

Organize the computer and software
Prepare the Tools and Working (Project) Environments
Basic issues with Master and Confidential Environments

TOOLS_2019.ZIP ^[2] contains FHOP macros and related files introduced in this volume

2. Standardizing Variables Over Time ^[3]

Time variables
Demographic variables
Confidential data elements

3. Preparing Master Files ^[4]

Setup activities
The RDYR macro
Check longitudinal consistency

4. Special Issues with Birth and Fetal Death Files ^[5]

Steps to make master files
Check longitudinal consistency
Geographic classification
Data quality

BC_FORMATS_2019.ZIP ^[6] contains FHOP's current format library for use with birth certificate and fetal death data (1989-2018)

11METH_WORK_2020.PDF ^[7] describes steps to clean work-related variables in the California 2010-2017 birth (mother and father) and 2014-2017 death (decedent) files and develop formats to classify those variables.

WORK_FORMATS_2020.ZIP [8] contains the resulting format library.

5. Special Issues with Death Files^[9]

Steps to make master files
Check longitudinal consistency
Cause of death
Geographic classification
Data quality

DT_FORMATS_2018.ZIP [10] contains FHOP's current format library for use with death certificate and fetal death data (1980-2018)

SD_GEOCODE7.PDF [11] summarizes work to evaluate the quality of address data in the California Death Statistical Master file in 2005 (before electronic death registration) and 2007 (after electronic death registration). It also compares the accuracy of two geocoding systems used in California at that time.

6. ^[12]Maintaining Hospital Formats^[13]

Structure of formats program
OSHPD facility labels
Centers for Medicare and Medicaid Services
Clinical Classification System (CCS)
Injury Classification

OSH_FORMATS_2019.ZIP [14] contains the SAS format library FHOP currently uses for OSHPD inpatient admissions (1983 to 2018) and emergency department and ambulatory care encounters (2005 to 2018). The files listed below are the source for the formats.

DXFH2018.ZIP [15] contains the last cross-classified lists of ICD-9 diagnoses (1983 to Sep-2015). This file is the source for formats that variously classify ICD-9 diagnosis codes

DXTFH2018.ZIP [16] contains the cross-classified lists of ICD-10 diagnoses (Oct-2015 - Dec-2017). This file is the source for formats that variously classify ICD-10 diagnosis codes

GEMI9I10.ZIP [17] contains the longitudinal crosswalk between the ICD-9 and ICD-10 diagnosis codes. This file is the source for formats that back-classify ICD9 to be consistent with current ICD-10 groupings.

ICD10_CONVERSION_2020.PDF [18] describes the work to validate the longitudinal GEMS crosswalk between the ICD-9 and the ICD-10 diagnosis codes with a focus on the Clinical Classification System, and particularly mental health (DXCH06) and conditions occurring during pregnancy, birth, and the puerperium (DXCH11).

PXAH2018.ZIP [19] contains the last cross-classified lists of ICD-9 procedures (1983 to Sep-2015). This file is the source for the formats that variously classify ICD-9 procedure codes. CCS did not update ICD-9 procedure codes in 2015.

PXTFH2018.ZIP [20] contains the cross-classified lists of ICD-10 procedures (Oct-2015 - Dec-2017). This file is the source for the formats that variously classify ICD-10 procedure codes

7.Maintaining Geography Formats[21]

The need for longitudinal geographic datasets
Standard administrative boundaries
Planning and policy geography
Data sets with geographic boundaries

GEOG_FORMAT_INPUT_2020.ZIP [22] contains the SAS programs and current input excel file used to make formats

GEOG_FORMATS_2020.ZIP [23] contains the full set of California geography formats in current use

8.Annual Hospital Disclosure Report[24]

Primary hospital data sets
Preparing AHDR data
Reconciling hospital events

9.Hospital Crosswalk[25]

Why crosswalk is needed
Crosswalk methods and results
Crosswalk validation
Example: Hospital-level race/ethnic data quality
The following files have longitudinal hospital-level results
Birth Certificate [26]
Patient Discharge [27]
Emergency Department [28]

10.Population Master Files [29]

Department of Finance
National Population Estimates
Intercensal Small Area Population Estimates

Issues and Decisions to be made on Collecting, Coding and Reporting Race and Ethnicity for Public Health Indicators [30]

The ?Race/Ethnicity Guidelines?, approved in 2003 by the California Directors of Public Health (CDPH) and Health and Human Services (CHHS) for use by all programs, explicitly did not address how to handle multi-race coding for trend analysis. Further, the National Center

for Health Statistics (NCHS) had not yet provided guidance on what to do when the same groups are not available over time or there is a mismatch between groups in the numerator and denominator. This document discusses issues related to developing a standardized approach to coding and reporting race and ethnicity for data sets maintained by CDPH. The focus is using these to explore race/ethnic differences in indicators of health status and outcomes over time. (September 2011).

Creating Longitudinal Hospital-Level Data Sets^[31]

Per California regulations, hospital licenses are based on a given physical location. When hospitals disappear from various data files the explanation is not readily apparent. We must determine whether it is because the facility closed, merged, converted to consolidated reporting, or moved, resulting in a new license ID. Yet another possibility is that a new license ID was assigned to a facility at the same location. We developed a series of decision rules to resolve such issues in a longitudinally consistent manner. These included rules to handle changes in hospital identifiers, physical location, consolidated data reporting, ownership, organizational type, and structural capacity. This document provides a full discussion of the issues encountered in creating the hospital-level data sets, their resolution, and the creation of related analysis data sets and variables. (June 2004)

Methods to Prepare Hospital Discharge Data^[32]

OSHPD distributes Patient Discharge Data (PDD) to qualified researchers such as the Family Health Outcomes Project (FHOP). The FHOP human subjects protocols permit us to have the confidential PDD, for all discharges and ages, from 1983 forward. Currently we have processed all years through 2000 and are about to start with the 2001 and 2002 files. This document presents an overview of the methods we developed to create the core files we use as the source for the different PDD-based research and data products that FHOP distributes. (June 2004)

Contact Us
UCSF Main Site

© 2013 The Regents of the University of California

Source URL: <https://fhop.ucsf.edu/data-management-methods>

Links

- [1] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/1METH_BASIC_COMPUTER_2018.pdf
- [2] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/TOOLS_2019_0.zip
- [3] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/2METH_STANDARD_VARIABLES_2021.pdf
- [4] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/3METH_MASTER_STANDARDS_2018.pdf
- [5] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/4METH_BIRTH_FDEATH_2018.pdf
- [6] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/BC_FORMATS_2019.zip
- [7] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/11METH_WORK_2020.pdf
- [8] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/WORK_FORMATS_2020.ZIP
- [9] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/5METH_DEATH_2018.pdf
- [10] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/DT_FORMATS_2018.zip
- [11] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/SD_GEOCODE7.pdf

- [12] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/6METH_OSHPD_FORMATS_2020.PDF
- [13] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/6METH_OSHPD_FORMATS_2020.pdf
- [14] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/OSH_FORMATS_2019.zip
- [15] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/DXFH2018.zip
- [16] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/DXTFH2018.zip
- [17] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/GEMI9110.zip
- [18] https://fhop.ucsf.edu/fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/ICD10_CONVERSION_2020.PDF
- [19] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/PXAH2018.zip
- [20] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/PXTFH2018.zip
- [21] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/7METH_GEOMAST_2020.pdf
- [22] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/GEOG_FORMAT_INPUT_2020.ZIP
- [23] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/GEOG_FORMATS_2020.zip
- [24] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/10METH_POP_2020.pdf
- [25] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/9METH_HOSP_XWALK.pdf
- [26] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/BRACEQH.XLS
- [27] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/PRACEQH.xls
- [28] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/ERACEQH.XLS
- [29] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/10METH_POP_2020B.pdf
- [30] http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/ISSUES_RACE_ETH_0.pdf
- [31] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/wysiwyg/METH_HOSP.pdf
- [32] https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/wysiwyg/METH_PDDS.pdf