

CREATING A LONGITUDINAL HEALTH FACILITIES DATABASE

Office of Health Care Statistics
Utah Department of Health

INTRODUCTION

Sterling Petersen, Analytics Lead
Office of Health Care Statistics
Utah Department of Health
sterlingpetersen@utah.gov

OFFICE OF HEALTH CARE STATISTICS (OHCS)

- Housed within the Center for Health Data and Informatics (CHDI) at the Utah Department of Health (UDOH)
- Collects, analyzes, and disseminates data
 - Healthcare Facility Discharge Data
 - Inpatient (1992–Current)
 - ED (1996–Current)
 - Ambulatory Surgery (1996–Current)
 - APCD (2013–Current)
 - HEDIS
 - CAHPS

HEALTHCARE FACILITIES DATABASE (HFD)

- Submitted on a quarterly basis, processed annually
 - Inpatient (1992-Current)
 - ED (1996-Current)
 - Ambulatory Surgery (1996-Current)
- Data for 146 Utah facilities
- Fully-identified discharge information
 - Name, address, DOB, sex
 - Diagnoses, procedures, e-codes
 - Discharge status
 - Payer information

HEALTHCARE FACILITIES DATABASE (HFD)

- Problems
 - Difficult to stack data from multiple years
 - Differing available variables
 - Differing variable names
 - Differing layouts (wide vs. long)
 - Difficult to detect readmission patterns
 - No longitudinal patient identifier
 - Decreasing enterprise use of SAS
 - Use of Microsoft SQL Server
 - Use of R
 - Use of Python

HEALTHCARE FACILITIES DATABASE (HFD)

- How to solve these issues?
 - Decide what variables are most important
 - Example: Is age really important to migrate?
 - Develop a common data format with the most important variables
 - Long format, one procedure/revenue code per line
 - Pipe-delimited
 - Determine final dataset for each year and each data type (IP, ED, and AS)
 - Challenge: Multiple datasets with ambiguous versioning
 - Inspect all variables in all years and identify content
 - Example: What does CI mean?
 - Determine best variables representing content
 - Example: PROC I vs. CPT I
 - Manipulate data into common data format

HEALTHCARE FACILITIES DATABASE (HFD)

- How to solve these issues?
 - Develop database schema, define tables
 - Header
 - Facility
 - Patient
 - Insured
 - Payer
 - Provider
 - Revenue
 - Load the data from the common data format file into the database schema
 - Quality checks
 - Repeat process as needed to correct errors

HEALTHCARE FACILITIES DATABASE (HFD)

- How to solve these issues?
 - Resolve duplicates in patient table
 - Identity resolution
 - Informatics Program
 - Utah Department of Health Master Person Index
 - dohMPI
 - Ongoing effort, currently underway

CONCLUSION AND LESSONS LEARNED

- Loss of institutional knowledge is bad
 - Document
 - File documentation in predictable places
- Stacking datasets can be very hard
 - Iterative process, learn as you go
- Ensuring data quality on original intake is best
 - Not feasible to get resubmissions for data collect more than a year ago
- Garbage in, garbage out
 - Incomplete or erroneous data is still incomplete or erroneous