



Putting the Pieces Together

Linked Data Products

Vital Statistics Death/PDD

Presenter:
Jan Morgan

osbpd



OSHPD Linked Data Sets

- **CCORP/PDD**

Patient Discharge Data linked to CABG clinical data collected by the Health Outcomes Center.

- **PDD/Vital Statistics Birth Cohort**

Maternal delivery and Infant birth records linked to vital statistics data

- **PDD/Vital Statistics Death**



Patient Discharge Data

- OSHPD collects, maintains, and disseminates patient discharge data records from California hospitals
 - ~3.5 million records
 - all non-federal hospitals
 - demographic, diagnostic, and utilization data
- Questions determine data needs.
- No single data set has all pertinent information regarding all questions.



Death Statistical Master File

- **Includes all deaths occurring within California borders**
 - Includes deaths of California residents that occur within other states
 - Excludes deaths of California residents that occur outside of the U.S.
 - Excludes non-California residents who die outside of California
- **Results are maintained and released by the Department of Public Health, Office of Vital Statistics**



Linking Discharge With Death

- A Measure of Patient Outcome
 - PDD reports hospital mortality, excludes out-of-hospital deaths
 - Overall mortality is a more valid measure - account for early discharge, transfer, etc.

- Allows Assessment of Readmission
 - Without accounting for out-of-hospital deaths, readmission rates will be underestimated



Deterministic and Probabilistic Linkage

- **Deterministic linkages** are based on exact matching of all merge variables
- **Probabilistic linkages** are based on exact matching of some merge variables (**blocking**) with scoring partial matches on the other merge variables



Deterministic Linkage Failure

1. SSN error in either DSMF or PDD
 - DSMF - death certificates transcribed
 - PDD - patient SSN incorrectly transcribed at first admission at a particular hospital
 - Use of SSN by spouse or other family member
2. SSN missing in either DSMF or PDD
 - unreported or invalid SSN



Potential Sources of Bias

- Unmatched in-hospital deaths were more likely to be from:
 - minorities (11.2% AfrA vs 8.0 % white)
 - women (10.5% vs 6.0 % for men)
 - younger (18% for young adults)
 - Certain DRGs
 - trauma, stroke, severe mental illness
 - public hospitals



The PDD-Death Data Product

- **Probabilistic linkage** between the PDD and Death Statistical Master File using available personal identifying and ancillary information in each data set
- Patient Discharge Database
 - SSN, gender, birth date, race/ethnicity, **Zip code**, expected source of payment, date of admission, **date of discharge**, **hospital death**, principle diagnosis
 - Hospital code, Hospital Zip code
- Death Statistical Master File
 - Name, SSN, gender, birth date, race/ethnicity, **Zip code**, **date of death**, location of death, cause of death.



The PDD-Death Data Product

- Each death reported in the DSMF is linked to exactly **one** discharge abstract from the PDD (the last identifiable hospitalization)
- Each file contains all hospitalizations in a given year where there was a subsequent death



Probabilistic Linkage - Methods

1. **Assume** - Unique identifiers have errors
2. **Find** - Other identifying information
3. **Create** - Potential matches via blocking
4. **Grade** - Matches via scoring algorithm
5. **Select** - Matches via selection score
6. **Correct** - Remove inconsistent matches



Vital Statistics Death/PDD

- For each year of deaths, certificate data is linked to the most recent hospitalization in California.

By DSMF Year

By PDD Year

Match DSMF to PDD



**Remove
Matched
Records**

**Unmatched DSMF records used for
matching to earlier years (“Remainder”)**



Vital Statistics Death/PDD

- Version A : Linked to last hospitalization with any “type of care”
- Version B : Linked to last “acute care” hospitalization
- There is one record for each matching hospitalization and death certificate



2005 Linkage Results

PDD Total Records

3,990,255



PDD "Clean" Records

2,108,173



DSMF Total Records

239,228



DSMF "Clean" Records

232,300



Linked Records

148,158



What data is available?

- Currently data for 1990 – 2005 is available.
- 1990 is a half year, beginning July 1. This is when OSHPD began collecting SSN.
- The linkage for 2006 and 2007 is currently in process.



Vital Statistics Death/PDD

- Data requests should be directed to the OSHPD – Healthcare Information Division (HID)
- Contact LOUISE HAND – OSHPD/HID/HIRC
- Telephone: (916) 326-3813
- E-mail: LHand@oshpd.ca.gov
- Website: www.oshpd.ca.gov (<http://www.oshpd.ca.gov/>)
- For web issues contact: OSHPDWebmaster@oshpd.ca.gov